

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
23 September 2004 (23.09.2004)

PCT

(10) International Publication Number
WO 2004/080148 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2003/030720

(22) International Filing Date:
30 September 2003 (30.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/416,186 2 October 2002 (02.10.2002) US

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/416,186 (CIP)
Filed on 22 October 2002 (22.10.2002)

(71) Applicant (for all designated States except US): NU-
VELO, INC. [US/US]; 675 Almanor Avenue, Sunnyvale,
CA 94085 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TANG, Y., Tom
[US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).
ASUNDI, Vinod [US/US]; 709 Foster City Blvd., Foster
City, CA 94404 (US). REN, Feiyan [US/US]; 7703 Oak
Meadow Court, Cupertino, CA 95014 (US). ZHANG, Jie
[CN/US]; 4930 Popular Terrace, Campbell, CA 95008
(US). WEHRMAN, Tom [US/US]; CCSR Mol Pharm
3210, 269 W. Campus Dr, Stanford, CA 94305 (US).
WANG, Zhiwei [CN/US]; 836 Alturas Ave #B36, Sun-
nyvale, CA 94085 (US). MA, Yunqing [CN/US]; 280 W.
California Ave #206, Sunnyvale, CA 94086 (US). WANG,
Dunrui [CN/US]; 12252 Pepper Tree Lane, Poway, CA
92064 (US). CHEN, Rui-hong [US/US]; 1031 Flying
Fish Street, Foster City, CA 94404 (US). ZHAO, Qing,

A. [CN/US]; 1556 Kooser Rd, San Jose, CA 95118 (US).
WANG, Jian-Rui [CN/US]; 744 Stendal Lane, Cupertino,
CA 95014 (US). GHOSH, Malabika [IN/US]; 620 Iris
Avenue, Apartment 133, Sunnyvale, CA 94086 (US).
XUE, Aidong, J. [US/US]; 1621 South Mary Avenue,
Sunnyvale, CA 94087 (US). WENG, Gezhi [US/US]; 415
Moraga Avenue, Piedmont, CA 94611 (US). ZHOU, Ping
[US/US]; 7595 Newcastle Drive, Cupertino, CA 95014
(US).

(74) Agents: QUERTERMOUS, Elena et al.; Nuvelo, Inc.,
670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT,
RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,
SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished
upon receipt of that report
- with sequence listing part of description published sepa-
rately in electronic form and available upon request from
the International Bureau

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and
uses thereof.



WO 2004/080148 A2

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Provisional Application Serial No. 60/416,186 filed October 2, 2002 entitled "Novel Nucleic Acids and Polypeptides", which contains material previously disclosed in the following applications: U.S. Application Serial No. 10/084,643 filed February 26, 2002 entitled "Novel Nucleic Acids and Polypeptides", Attorney Docket No. 21272-502; PCT Application Serial No. PCT/US00/35017 filed December 22, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP3A/PCT; PCT Application Serial No. PCT/US01/02623 filed January 25, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785CIP3/PCT; PCT Application Serial No. PCT/US01/03800 filed February 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP3/PCT; PCT Application Serial No. PCT/US01/04927 filed February 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP3/PCT; PCT Application Serial No. PCT/US01/04941 filed March 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789CIP3/PCT; PCT Application Serial No. PCT/US01/08631 filed March 30, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 790CIP3/PCT; PCT Application Serial No. PCT/US01/08656 filed April 18, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 791CIP3/PCT; all of which are incorporated herein by reference in their entirety.

2. BACKGROUND OF THE INVENTION

2.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, circulating soluble factors, chemokines, and interleukins) has

matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These

nucleic acid sequences are designated as SEQ ID NO: 1-684, or 1369-1966 and are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases or unknown. In the amino acids provided in the Sequence Listing, an asterisk (*) corresponds to the stop codon.

5 The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1-684, or 1369-1966 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ
10 ID NO: 1-684, or 1369-1966. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1-684, or 1369-1966 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

 The nucleic acid sequences of the present invention also include the sequence
15 information from the nucleic acid sequences of SEQ ID NO: 1-684, or 1369-1966. The sequence information can be a segment of any one of SEQ ID NO: 1-684, or 1369-1966 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-684, or 1369-1966.

 A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be
20 provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

 This invention also includes the reverse or direct complement of any of the nucleic acid
25 sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media,
30 use in sequencing full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-684, or 1369-1966 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-684, or 1369-1966 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., *Science* 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-684, or 1369-1966; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-684, or 1369-1966; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-684, or 1369-1966. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-684, or 1369-1966; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in SEQ ID NO: 1-684, or 1369-1966; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homologue (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in SEQ ID NO: 685-1368, or 1967-2564, or Tables 3A, 3B, 5, 7, or 8.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-684, or 1369-1966; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

5 The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in
10 which the protein produced by such processes is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in
15 generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, e.g., *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as
20 expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a
25 polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical
30 condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

5 The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a
10 complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under
15 conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs,
20 and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the
25 identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell
30 for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can affect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Tables 2A and 2B); for which they have a signature region (as set forth in Tables 3A and 3B); or for which they have homology to a gene family (as set forth in Tables 4A and 4B). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms “a”, “an” and “the” include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms “biologically active” or “biological activity” refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise “immunologically active” or “immunological activity” refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms “complementary” or “complementarity” refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded

molecules may be "partial" such that only certain portion(s) of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

5 The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell
10 lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are
15 able to regenerate themselves.

 The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

 As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF.

20 EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

 The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or
25 "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and
30 N is A, C, G, or T (U) or unknown. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of

oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NO: 1-684, or 1369-1966.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-684, or 1369-1966. The sequence information can be a segment of any one of SEQ ID NO: 1-684, or 1369-1966 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-684, or 1369-1966, or those segments identified in Tables 3A, 3B, 5, 7, or 8. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-

mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes.

Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1/4^{25}$) times the increased probability for mismatch at each nucleotide position (3×25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids.

Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full-length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant" (or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, e.g., recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be

reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (e.g., microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or

elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

5 The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that
10 are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (*e.g.* Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2): 134 -143) and factors released from damaged cells (*e.g.* Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

15 Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

 The term "stringent" is used to refer to conditions that are commonly understood in
20 the art as stringent. Stringent conditions can include highly stringent conditions (*i.e.*, hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (*i.e.*, washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

25 In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

 As used herein, "substantially equivalent" or "substantially similar" can refer both to
30 nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of

those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have

5 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than

10 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more than 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least

15 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at

20 least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of

25 determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a new stop codon) should be disregarded. Sequence identity may be determined, *e.g.*, using the Jotun Hein method (Hein, J. (1990) *Methods Enzymol.* 183:626-645). Identity between sequences can also be determined by other methods known in the art, *e.g.* by varying hybridization conditions.

30 The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal

integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

5 As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid
10 molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

15

4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-684, or 1369-1966; a polynucleotide encoding
20 any one of the peptide sequences of SEQ ID NO: 1-684, or 1369-1966; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-684, or 1369-1966. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID
25 NO: 1-684, or 1369-1966; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing, or Table 7; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homologue of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO:
30 685-1368, or 1967-2564 (for example, as set forth in Tables 3A, 3B, 5, 7, or 8). Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable

immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

5 The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

10 The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 15 1-684, or 1369-1966 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-684, or 1369-1966 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-684, or 1369-1966 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

20 The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

25 The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, e.g., at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, 30 and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide

sequences of SEQ ID NO: 1-684, or 1369-1966, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-684, or 1369-1966, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-684, or 1369-1966 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology results for the nucleic acids of the present invention, including SEQ ID NO: 1-684, or 1369-1966 can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST (Basic Local Alignment Search Tool) program is used to search for local sequence alignments (Altschul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using FASTXY algorithm may be performed.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be

prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA

fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention could be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-684, or 1369-1966, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient

restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

5 The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-684, or 1369-1966 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-
10 684, or 1369-1966 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present
15 invention. The following vectors are provided by way of example: Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

20 The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman,
25 *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

30 Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate

early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the
5 ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with
10 translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for
15 bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for
20 transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from
25 commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain
30 and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation,

disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., Nat. Biotech 17, 870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-684, or 1369-1966, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 1-684, or 1369-1966 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1-684, or 1369-1966 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1-684, or 1369-1966, antisense nucleic acids of the invention can be designed

according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (*v*), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (*acp3*)*w*, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of

the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1-684, or 1369-1966). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA. See, *e.g.*,

Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

5 Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

10 In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose
15 phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis
20 protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

 PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting
25 replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

30 In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may

combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of
5 base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine
10 phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5:
15 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication
20 No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport
25 agent, a hybridization-triggered cleavage agent, etc.

4.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids
30 of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are

in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and

eukaryotic hosts are described by Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express
5 recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal
10 diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived
15 from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used,
20 as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as
25 yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or
30 bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene
5 targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, and regulatory protein binding sites or combinations of said sequences.
10 Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA
15 molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory
20 element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are
25 contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting
30 sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. 5 PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a
10 polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 685-1368, or 1967-2564 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-684, or 1369-1966 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the
15 nucleotide sequences set forth in SEQ ID NO: 1-684, or 1369-1966 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 685-1368, or 1967-2564 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as
20 SEQ ID NO: 685-1368, or 1967-2564 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological
25 activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 685-1368, or 1967-2564.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein
30 may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as

immunoglobulins for many purposes, including increasing the valency of protein binding sites. Fragments are also identified in Tables 3A, 3B, 5, 7, or 8.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein
5 coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The predicted signal sequence is set forth in Table 5. The mature form of such protein may be obtained and confirmed by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell and sequencing of the cleaved product. One of skill in the art will recognize that the actual cleavage site may be different
10 than that predicted in Table 5. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed (See, e.g.,
15 Sakal et al., Prep. Biochem. Biotechnol. (2000), 30(2), pp. 107-23, incorporated herein by reference).

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic
20 acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the
25 ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or
30 tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may

be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

5 The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic
10 sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the
15 methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments
20 include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated
25 polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include
30 fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then
5 tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that
10 are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 685-1368, or 1967-2564.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the
15 protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of
20 interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S.
25 Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for
30 biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

5 The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described
10 in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

 The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting
15 expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™;
20 one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

 Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as
25 a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

30 Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide

a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces
5 fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability.
10 Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic
15 agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE

20 IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group,
25 University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by
30 reference), Pfam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobicity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), the GeneAtlas software (Molecular Simulations Inc. (MSI), San Diego, CA) (Sanchez and Sali (1998) Proc. Natl. Acad. Sci., 95,

13597-13602; Kitson DH et al, (2000) "Remote homology detection using structural modeling – an evaluation" Submitted; Fischer and Eisenberg (1996) Protein Sci. 5, 947-955), Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark) incorporated herein by reference).

5 Polypeptide sequences were examined by a proprietary algorithm, SeqLoc that separates the proteins into three sets of locales: intracellular, membrane, or secreted. This prediction is based upon three characteristics of each polypeptide, including percentage of cysteine residues, Kyte-Doolittle scores for the first 20 amino acids of each protein, and Kyte-Doolittle scores to calculate the longest hydrophobic stretch of the said protein. Values of
10 predicted proteins are compared against the values from a set of 592 proteins of known cellular localization from the Swissprot database (<http://www.expasy.ch/sprot>). Predictions are based upon the maximum likelihood estimation.

Pesence of transmembrane region(s) was detected using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html).

15 The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.7 CHIMERIC AND FUSION PROTEINS

20 The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the
25 invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

30 For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

5 In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in*
10 *vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a
15 subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional
20 techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of
25 gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety
30 (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be

modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences.

5 See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with
10 the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control
15 of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment
20 regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion
25 properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple
30 deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are

deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by
5 supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to
10 express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or
15 inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals,
20 can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No
25 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or
30 even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment.

Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA

sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies
5 using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the
10 other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or
15 its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

20 Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F.
25 Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as
30 nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid

preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

5 **4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION
ACTIVITY**

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations.

10 A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell
15 proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those
20 described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology
25 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan
30 eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin- γ , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors.

The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotent/pluripotent stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotent/pluripotent mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies

would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., *Differentiation*, 48: 173-182, (1991); Klug et al., *J. Clin. Invest.*, 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. *Proc. Natl. Acad. Sci. U.S.A.*, 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the

invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

5 A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, 10 thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in 15 supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and 20 therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or 25 heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

30 Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994;

5 HIRAYAMA et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In *Culture of Hematopoietic Cells*.

10 R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc.,

15 New York, N.Y. 1994.

4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and

20 tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have

25 prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming

30 cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast

activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from

chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with
5 vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising
10 such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and
15 conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

20 Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in:
25 Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

30 A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and

disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also be useful in the treatment of allergic reactions and conditions (e.g., anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animal models such as the cumulative contact enhancement test (Lastbom et al., *Toxicology* 125: 59-66, 1998), skin prick test (Hoffmann et al., *Allergy* 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., *Arch. Toxicol.* 73: 501-9), and murine local lymph node assay (Kimber et al., *J. Toxicol. Environ. Health* 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of

an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing
5 non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without
10 limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition
15 as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may
20 avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in
25 humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed.,
30 Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self-tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected
5 with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g.,
10 B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T
15 cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation,
20 those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J.
25 Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching
30 (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro

antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., *Endocrinology* 91:562-572, 1972; Ling et al., *Nature* 321:779-782, 1986; Vale et al., *Nature* 321:776-779, 1986; Mason et al., *Nature* 318:659-663, 1985; Forage et al., *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986.

4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to

tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostasis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis

Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.10.11 CANCER DIAGNOSIS AND THERAPY

5 Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing
10 malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation,
15 inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies
20 including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal
25 neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central
30 nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Kaposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention

(including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987)

Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis
5 assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

10

4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such
15 receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and
20 humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured
25 by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions
30 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

5 Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182
10 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

15 **4.10.13 DRUG SCREENING**

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One
20 method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or
25 fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries
30 comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product
5 libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide
10 and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol.* 9(3):205-23 (1998); Hruby
15 et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then
20 tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The
25 toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.10.14 ASSAY FOR RECEPTOR ACTIVITY

30 The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening

assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The responses of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications *i.e.* phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an

inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflammation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic myelogenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include

but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

- (i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or
5 compression injuries;
- (ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;
- (iii) infectious lesions, in which a portion of the nervous system is destroyed or
10 injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;
- (iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration
15 associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;
- (v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency,
20 Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;
- (vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;
- (vii) lesions caused by toxic substances including alcohol, lead, or particular
25 neurotoxins; and
- (viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various
30 etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival

or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or *in vivo*;
- 5 (iii) increased production of a neuron-associated molecule in culture or *in vivo*,
e.g., choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method
10 set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons
may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or
Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of
neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody
binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor
15 neuron dysfunction may be measured by assessing the physical manifestation of motor
neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the
invention include but are not limited to disorders such as infarction, infection, exposure to
toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor
20 neurons as well as other components of the nervous system, as well as disorders that
selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited
to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis,
infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-
Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory
25 Neuropathy (Charcot-Marie-Tooth Disease).

4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following
additional activities or effects: inhibiting the growth, infection or function of, or killing,
30 infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites;
effecting (suppressing or enhancing) bodily characteristics, including, without limitation,
height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or
organ or body part size or shape (such as, for example, breast augmentation or diminution,

change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s);
5 effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of
10 the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

15

4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential
20 predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in
25 humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate
30 fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that

hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with
5 nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

10 Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.10.20 ARTHRITIS AND INFLAMMATION

15 The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et al., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single
20 injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

25 The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would
30 reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

5

4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An
10 exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of
15 polypeptide administered per dose will be in the range of about 0.01 µg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1 µg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution,
20 dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

25 4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be
30 administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other

materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound

sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient
5 alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention
10 is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other
15 active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic
20 factor(s), thrombolytic or anti-thrombotic factors.

4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular,
25 subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous,
30 intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in

fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

15

4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical

30

composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from
5 about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally
10 acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride
15 Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or
20 physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such
25 carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable
30 excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired,

disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such

as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable

matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides,

diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like.

Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

5 The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient.

10 Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to

15 practice the method of the present invention should contain about 0.01 μ g to about 100 mg (preferably about 0.1 μ g to about 10 mg, more preferably about 0.1 μ g to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition

20 topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage.

 Topical administration may be suitable for wound healing and tissue repair. Therapeutically

25 useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active

30 ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, 5 tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised 10 of combinations of any of the above-mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having 15 diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, 20 ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful 25 herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, 30 proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet

derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

25

4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be

30

estimated initially from appropriate *in vitro* assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as
5 determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic
10 efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50} . Compounds which exhibit high therapeutic
15 indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of
20 administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in*
25 *vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of
30 the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 $\mu\text{g/kg}$ to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 $\mu\text{g/kg}$ to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at
5 longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

10 4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a
15 compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain,
25 F_{ab} , F_{ab}' and $\text{F}_{(\text{ab})_2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a
30 reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 685-1368, or 1967-2564, or Tables 3A, 3B, 5, 7, or 8, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine

binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), *Antibodies A Laboratory Manual*; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known

in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8
5 (April 17, 2000), pp. 25-28).

4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular
10 species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

15 Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256, 495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be
20 immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell
25 line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that
30 preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107, 220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as

a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA
5 also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted
10 for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

4.13.3 HUMANIZED ANTIBODIES

15 The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab',
20 F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321, 522-525 (1986); Riechmann et al., Nature, 332, 323-327 (1988); Verhoeyen et al., Science, 239, 1534-1536 (1988)), by substituting
25 rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise
30 substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion

of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2, 593-596 (1992)).

5 4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human
10 B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80,
15 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227, 381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by
20 introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806;
25 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368, 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13, 65-93 (1995)).

30 Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains

in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then
5 obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after
10 immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for
15 example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent
20 rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

25 A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The
30 hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

4.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

5 According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246, 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or
10 derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab)2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab)2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing
15 agent and (iv) F_v fragments.

4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of
20 the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two
25 immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305, 537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished
30 by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10, 3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121, 210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229, 81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab' -TNB derivatives is then reconverted to the Fab' -thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab' -TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175, 217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical

coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5), 1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90, 6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152, 5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147, 60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG ($Fc\gamma R$), such as $Fc\gamma RI$ (CD64), $Fc\gamma RII$ (CD32) and $Fc\gamma RIII$ (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176, 1191-1195 (1992) and Shopes, J. Immunol., 148, 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53, 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3, 219-230 (1989).

4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolacca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

25

4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the

30

presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-684, or 1369-1966 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-684, or 1369-1966 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein-encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the

present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means
5 having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access
10 manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target
15 sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available
20 algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The
25 most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally
30 selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include,

but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

4.15 TRIPLE HELIX FORMATION

5 In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple
10 helix-see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 15241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization
15 blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

20 4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

25 In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization
30 conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

5 In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods
10 employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science
15 Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or
20 membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is
25 compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies
30 of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-684, or 1369-1966, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

- (a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and
5 detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting
10 the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression
15 of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to
20 activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in
25 the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

30 For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed"

when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al.,
5 Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or
10 EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple
15 helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix -
20 see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Okano, J. Neurochem. 56, 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks
25 translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention
30 can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-684, or 1369-1966. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NO: 1-684, or 1369-1966 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well-known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal

map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

5 4.20 **PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES**

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those
10 of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6), 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988;
15 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8), 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be
20 purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the
25 microwell surface termed CovaLink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridgeheads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

30 The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen *et al.*, (1991). In this technology, a phosphoramidate bond is employed (Chu *et al.*, (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins

the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently
5 bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ μ l) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm₇), is then added to a final concentration of 10 mM 1-MeIm₇. A ss DNA solution is then dispensed into CovaLink NH strips (75 μ l/well) standing on ice.
10 Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm₇, is made fresh and 25 μ l added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS
15 heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link
20 to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe
25 arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995), 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res., 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988)
30 Anal. Biochem. 169(1), 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) Proc. Nat'l. Acad. Sci., USA 91(11), 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24), 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*II normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*II**), yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992), quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*II** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*II** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 µg instead of 2-5 µg); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm², depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be

spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm² and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5.0 EXAMPLES

5.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences.

5 **5.2 EXAMPLE 2**

Assemblage of Novel Nucleic Acids

 The contigs or nucleic acids of the present invention, designated as SEQ ID NO: 1369-1966 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from
10 different databases (i.e., Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene, and exons from public domain genomic sequences predicated by GenScan) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the above databases that would extend the assemblage. Further, inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with
15 BLAST score greater than 300 and percent identity greater than 95%.

 Table 7 sets forth the novel predicted polypeptides (including proteins), SEQ ID NO: 1967-2564, encoded by the novel polynucleotides (SEQ ID NO: 1369-1966) of the present invention, and their corresponding translation start and stop nucleotide locations to each of SEQ ID NO: 1369-1966. Table 7 also indicates the method by which the polypeptide was predicted.
20 Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of the translated novel polynucleotide to known polynucleotides (W.R. Pearson, Methods in Enzymology, 183:63-98 (1990), herein incorporated by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for
25 human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268:78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary
30 strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame.

5.3 EXAMPLE 3

Novel Nucleic Acids

The novel nucleic acids of the present invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full-length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequences were checked using FASTY and/or BLAST against Genbank (i.e., dbEST, gb pri, UniGene, and Genpept) and the Geneseq (Derwent). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NO: 1-1368.

The nucleic acid sequences of the present invention were confirmed to have at least one transmembrane domain using the TMPred program (http://www.ch.embnet.org/software/TMPRED_form.html). One of skill in the art will recognize that the proteins of the present invention may be utilized as either a membrane-bound target or a soluble protein.

Table 1 shows the various tissue sources of SEQ ID NO: 1-684.

The homologs for polypeptides SEQ ID NO: 685-1368 that correspond to nucleotide sequences SEQ ID NO: 1-684 were obtained by a BLASTP version 2.0a1 19MP-WashU searches against Genpept and Geneseq (Derwent) using BLAST algorithm. The results showing homologues for SEQ ID NO: 685-1368 are shown in Tables 2A and 2B.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6, 219-235 (1999), <http://motif.stanford.edu/ematrix-search/> herein

incorporated by reference), all the polypeptide sequences were examined to determine whether they had identifiable signature regions. Scoring matrices of the eMatrix software package are derived from the BLOCKS, PRINTS, PFAM, PRODOM, and DOMO databases. Tables 3A and 3B show the accession number of the homologous eMatrix signature found in the indicated polypeptide sequence, its description, and the results obtained which include accession number subtype; raw score; p-value; and the position of signature in amino acid sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Tables 4A and 4B show the name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence. Further description of the Pfam models can be found at <http://pfam.wustl.edu/>.

Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al reference, was obtained for the polypeptide sequences.

Table 6 correlates nucleotide sequences of the invention to a specific chromosomal location when assignable.

Table 8 shows the number of transmembrane regions, their location(s), and TMPred score obtained, for each of the SEQ ID NO: 685-1368 that had a TMPred score of 500 or greater, using the TMpred program (http://www.ch.embnet.org/software/TMPRED_form.html).

Table 9 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-684, their corresponding polypeptide sequences SEQ ID NO: 685-1368, their corresponding priority contig nucleotide sequences SEQ ID NO: 1369-1966, their corresponding priority contig polypeptide sequences SEQ ID NO: 1967-2564, and the US serial number of the

priority application (all of which are herein incorporated in their entirety), in which the contig sequence was filed.

114

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
adult brain	GIBCO	AB3001	39-40 56 68 93 154-155 189 205 215 221 229 245 289-290 296 298 305 307 314 324 346 362 376 384 438 444 493 499 502 532 563 612 624 654 668
adult brain	GIBCO	ABD003	10 13 15 17-20 27 29 34 40 47-49 56 61-63 66 68 75 80- 82 86 93-94 96 98 102 106 137 150 154 156-159 161 168-169 173-174 179 188 205 210 212 215 221 229- 231 243 245 290 296 302 305 307 313-315 319-320 323 325 331 346 349 352 359 362 367 371 376 384 420-421 428 438 444 447 461-462 473-474 487 493 499 516 519 522- 523 529 532 541 550 563 587-588 601 612 616 624 627 635 643 652- 654 660 669 672- 673 677-678
adult brain	Clontech	ABR001	7 18 22 24 29 47- 50 56 68 70 75 79 112-113 152 161 186 205-206 212 220 230 259-262 280 282 296 302 346 361 376 384 420 465 488-489 492 518 520 587 595 620-621 652 660 682
adult brain	Clontech	ABR006	7-8 10 13 16 20-21 23 27 34 37 40 53 56 64-65 69-70 73- 74 79 88-89 92 100 104-105 147-150 160-161 170 186 200 207 212 229- 230 243 256 259- 262 266 275-278 280 282-283 287 289-290 307 309 314-315 317-318 321-322 325 337-

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			338 349-352 357 359-360 364 377 384 430 447-448 461 466 484 499 501 503 518 520 530 532 542-546 552 556 562-563 569-571 600 607- 616 620-621 623- 625 628-629 641- 642 653 660 672- 673 677-678 682
adult brain	Clontech	ABR008	7-8 10 14 19 21 23 25-28 30-33 37-39 43 46-50 52-53 56- 57 59 62-65 67-68 73-76 86-89 92-94 104-105 118 131- 134 139-140 144 147-148 150 153- 154 160-165 170 180 186 189 205- 206 208-212 218- 219 223 229-230 232-234 236 242- 245 249 259-263 266 268 270 273 283-289 293 298 302 305 307-308 313-316 318-324 334-335 337-341 343 346 349 351 356 359 361-364 367 371 377 381 384 387-388 390 403-404 419 423- 425 431 435-436 438 440-441 445- 451 462 473-475 484 493 498-501 504-506 509 512 514-522 525 527 529-530 532 534 543-545 550 558 562-564 569 576 583-584 591 597- 599 601-602 605 607-610 620-621 624-625 627-628 631-632 638-640 652-653 660 663 665 670-671
adult brain	Clontech	ABR011	289 384 537
adult brain	BioChain	ABR012	26 384 607
adult brain	BioChain	ABR013	20 79 153 220 289 384 465 526
adult brain	Invitrogen	ABR014	48-50 52 106 170 230 335 384 430

116
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			438 501 530 536 635 643
adult brain	Invitrogen	ABR015	20 46 106 150 153 216 371 384 401 461 526 643
adult brain	Invitrogen	ABR016	60 69 153 368 384- 385 507 522 587 654
adult brain	Invitrogen	ABT004	10 16 24 29 43 47- 49 56 60 64 67-69 73 79 97-98 165 168-170 179 186 189 205 230 242- 247 249 259-263 289-290 296 298 305 308-310 314- 315 319 329-330 332-333 349 359 380 384-385 387- 388 390 428 451 456-457 475 487- 490 492-493 499- 500 512 519-520 522 529-530 587 612 620-621 643 654 663 665
cultured preadipocytes	Stratagene	ADP001	10 19-20 23 26 36 68 70 106 116-117 147-148 165 171- 172 189 220 246- 247 256 273 289 305 316-319 329- 330 349 351 361 365 392 394-398 400 423-424 428 451 465 487 499 507 522 529 534 543 587 643 672- 673 682
adrenal gland	Clontech	ADR002	10 18 25 27 29 47- 49 52-53 56 64 73- 75 83 87 90 100 106 110 124 130 137 144 160-161 163 182 189 198 200 202-203 208 211-212 215 217 220 237-241 249 251 259-263 280 289-293 296 317- 319 329-331 344- 345 359 362 371 377 384 390 403- 404 423-424 426 465 499-501 507 516 522 525 539 570 572-573 585 600-601 611 620-

117
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			621 623-624 635 643 660 663 672- 673 675
adult heart	GIBCO	AHR001	5 16 18 24-26 34 37 39 46 56 64 66- 68 75 77 83 86-89 92 94-97 101-102 104-106 110 134 150 154 158-159 162 168-170 194- 196 202-203 212 215 224-226 229 269 289 296 302 306 308-309 314 320 323-324 331 336-338 342 346 356 367 371 377- 378 384-385 390 400 402 417-418 421 428 431 436 438 447 461-462 475 479 484-485 491 498 501 507 516 518 522-525 530 532 534 541 554 564 570 572- 573 586-587 601 605 607 610 613- 614 635 643 652 662 669 672-673
adult kidney	GIBCO	AKD001	5 10 12-13 16 18 20 24-26 29 39 43 52 54 56 62-64 66 68 71-72 75-76 83 89-96 98 106-109 112-114 116-117 122-126 131 137 139 155 158-159 162 170 172-174 177 183-184 188 200 202-203 205 208 215-216 218- 219 229-230 245 247 256 268 272 275-278 289-290 296 298-299 302 308-309 314 316 319-320 323 329- 330 332-333 336 350 359-360 364 367-368 371 377 384 392-393 400 402 420 423-424 428 431 435-436 438 444 451 461 473-474 484-486 492-493 499-500 504-507 510 516

118
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			518-519 521-522 524 526 529-530 532 534 537 539 541 567-568 587- 588 613 620-621 623 631-632 635 643 652 654 664 668 672-673
adult kidney	Invitrogen	AKT002	6 8 10 14-15 17 20 24-25 29 33-34 40 46-50 64 67 75 80- 82 85 88 93-94 106 116-117 126 150 154 157 162-164 168-169 188 199 216-219 222 232- 234 255-256 271 275-278 289 296 298 308 312 317- 319 332-333 337- 338 348 358 360 368 370-371 384 390 400 421 430 435 438 451 461- 462 491-493 499- 501 507 509 516 518 520 522 524 530 535-537 552 564 567-568 580 587 597-599 607 631-632 635 643 652 662 666 669 672-673 675 677- 679
adult lung	GIBCO	ALG001	13 22 26 63 66 68 75 93 106 112-114 127-130 137 144 150 165 177 230 256 271 289 302 314 323 327 337 342-343 368 371 384 390 392-393 421 484 488-489 504-507 539 564 638-639 643 661 675
lymph node	Clontech	ALN001	13 26 33 54 56 128-131 135 150 166 173-174 202- 203 211 215-216 256 259-262 289 320 327 350 367- 368 371 465 507 509 526 643 669
young liver	GIBCO	ALV001	5 10 13 24-25 43- 44 56 67-68 71 80- 82 89 106 110-111 132-133 137 154

119
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			168-170 179 183- 184 205 218-219 221 229 275-278 296 302 320 367 371 390 428 438 487-490 498 502 507 525 530 538 635 641-643 651 666
adult liver	Invitrogen	ALV002	5 14 16-17 19 24- 25 37 52 64 66 68 80-82 87 90 93 97- 98 104-105 132-133 137 140 150 170 183 186 188 215 218-220 229 232- 234 249 256 272 275-278 289 294- 295 311-312 314 319 332-333 351 358-359 364 366 371 377 381 386- 387 392-393 428 449 451 465 487- 489 495-498 518 522 538 593 601 607 610 631-632 643 666
adult liver	Clontech	ALV003	7 18-19 24 38 46 180 186 216 220 222 249 275-278 371 390 427 465 495 499 530 538 623 627 632 666 679-680
adult ovary	Invitrogen	AOV001	5 7-8 10 12 14 16 18 20 25-27 29 33 36 38-40 47-49 53- 54 56 59 61-62 64 67-68 73-76 79-83 87 89 92-94 96 98 106-107 111-114 116-118 121 128- 131 134-135 137 139-142 150 153- 154 157-161 171- 177 179-180 182 187 189 194-198 200 202-203 205- 206 211 218-219 222 229-230 235- 241 245 249 251 254-256 259-264 267 272 282 289- 290 296 298-299 302 305-306 308 311-314 316 320 323-325 327 331-

120
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			333 336 342 346- 347 349-351 358 362 367-368 371 377 380 383-384 390 392-393 400 402 420-423 425 427-428 435-436 438 444 451 454 459-462 471 473- 474 484 487-489 491 493 498-499 501-502 504-507 511 516 518 521- 522 524 530 532 539 543 547-550 555-556 564-565 581 587 593 595 602 605 607 616 620-621 623-624 631-632 635 643 652-654 660 667- 669 679-680
adult placenta	Clontech	APL001	1-4 63-64 66 143 145-146 178 211 216 289 296 323 351 384 537 630
placenta	Invitrogen	APL002	1-4 7 51 68 85 98 151-152 192 208 215 256 259-262 305 319 332-333 384 428 499 533 602 627 654 666
adult spleen	GIBCO	ASP001	7 13-14 17 26 32 52 54 56 63 75 89 106 109 112-115 120 135 137 141- 142 144 154 157 173-174 179-180 186 205 208 216 220-222 229 252 256 259-262 272 279 289 296 298 302 308 312 319- 320 337-338 347 364 367-368 371 384 400 427 438 451 459-461 465 484 487 500 504- 507 522 525-526 530 534 555 587 593 617-618 631- 633 635 638-639 643 663 669 675- 676 679
adult testis	GIBCO	ATS001	5 10 19 29 39 64 68 93 100 106 116- 117 137 145-146 150 153 172 175-

121
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			176 181-182 198 202-203 229 249 256 267 289 296 298 302 305 307- 308 314 316 323 331 356 359 362 364 371 384 402 426 438 451 485 500 507 518-519 591 597-599 619- 621 643 654 662
adult bladder	Invitrogen	BLD001	5 10 26 51 65 68 84 89 93 131 175- 176 211 256 259- 262 267 289 314 317-318 332-333 351 383-384 395- 398 423-424 426 499 501 522 525 580 593 643 661 682
bone marrow	Clontech	BMD001	5 7 30-31 34 37 40 47-49 54-56 62 68 75-80 83 93 96 100 131 136 147-148 150 158-159 163 165 172 177 198 204 206 211 216 229 289 302 308 316 319-320 324- 325 337-338 350 358 364 367-368 371 400 422 428 438 452 454 461 478 484 487 491 499-502 507 509- 510 520 530 536- 537 541 543 554 587 624 638-639 643 651-652 654 667-669 672-673
bone marrow	GF	BMD002	7-8 12 14 17 20 25 27-28 32-33 37 43 52 57 63-64 66-68 77 87 100 102 106- 107 112-114 116- 118 120 131 136- 137 144 147-148 150-153 157-159 163 172 179 199 206 215-216 222 256 259-263 268 272 275-278 286 289 298 302-303 305 308 317-318 325 337-338 341 343 347-348 368 371 390 400 427-

122
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			428 430-431 434- 435 437 444 451 461-462 488-489 491-492 499 501- 502 504-507 509 511 516 520 525- 526 530 537 543 554 558 560-561 585 587 595 600 610 623 629 631- 633 635 638-640 643 667-669 672- 673 679
bone marrow	Clontech	BMD004	507 522
bone marrow	Clontech	BMD007	368 504-506 672
*Mixture of 16 tissues - mRNA	Various Vendors	CGd010	99 132-133 165 237-241 275-278 290 298 306 336 368 380 402 423- 424 509 556 586 610
*Mixture of 16 tissues - mRNA	Various Vendors	CGd011	33 42 153 168-169 178 213-214 245 247 467 526 537 572-573 675
*Mixture of 16 tissues - mRNA	Various Vendors	CGd012	5 14 18 21 24 31 33 35 39 42 44 46 51 53 58 61-62 70- 72 75 80 84-85 90 92-93 96 98 100- 103 127 131 144- 146 153-154 157 160-161 163 165 168-169 175-176 178-179 183 185 189 193 200 218- 219 221 229 232- 234 245 247 256 259-262 275-278 280 289-292 298 300-301 308 311 317-318 325 335- 338 342 344-347 349 352 355-356 359-360 368 370- 375 380 384-386 388 391 394-399 401-402 405-407 410 412-413 419 428 450-451 464 467-469 471 504- 507 512 516 518 524 526 532 537 541 545 547-549 554 556 563-564 572-573 586 590- 591 600 602 605 623-625 627-628

123

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			652 654 659-660 664 667 670-671 676 682
*Mixture of 16 tissues - mRNA	Various Vendors	CGd013	56 58 61-62 70 131 160-161 163-164 193 247 290 311 345 348 360 368 370 394-398 512 537 556 660 682
*Mixture of 16 tissues - mRNA	Various Vendors	CGd015	1-5 8 14 17 52 59 68 87 215 228 259- 262 272 275-278 289 309 371 377 392-393 400 402 420 446-447 451 492 498 504-506 514 521 537-538 588 620-621 637 643 654 672-675
*Mixture of 16 tissues - mRNA	Various Vendors	CGd016	10 14 19 24-28 33 57 65 70 76 112- 114 121 131 151- 153 163 183 206 218-219 325 328 332-333 394-398 435 440-441 488- 489 500 510 518- 520 532 569 590 641-643 653 662- 663 668 671-673 682
adult colon	Invitrogen	CLN001	5 10 14 29 35 47- 50 56 112-114 135 175-176 179 220 230 254 256 289- 290 308 332-333 343 368 371 385- 386 415 427-428 436 465 498 510 518 534 572-573 580 597-599 607 643 651 661 663 669
adult cervix	BioChain	CVX001	7 10 14 16 18 20 23-26 30-31 40 47- 49 56 62 66 70 73- 76 83 85 87 89 93- 94 97 103 106 126 131 137 141-142 144 147-148 154 175 177 179 182 188-189 197-198 202-203 206 211 221 229 245 249 259-263 267 282 287 289 296 298 302 305 308 314 320 323-325 329-

124
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			333 350 356 358 362 367-368 371 377 382 384 390 400 438 451 454 459-460 462 465 484 487-490 492- 493 499-502 507 516 522 524-525 530 532 534-535 541 550 555 572- 573 580 587 602 605 610 613-614 616 623-624 626 628 643 652 661 663-664 668 680 682
diaphragm	BioChain	DIA002	93 134 308 402
endothelial cells	Stratagene	EDT001	7 10 12 17 19 23 29 34 36 39 52 54 56 63-64 66 68 75 80-84 86-89 92-93 95-97 106-107 116- 117 127 131 137 139 147-148 150 154 157-159 168- 169 172 179 182 192 198-199 202- 203 208 211 215 217 220-221 230- 234 249 254 256 259-262 264 270 272 289-290 296 298 313-314 316 320 323-324 348- 350 364 367 371 376-377 390 392 430 435 438 445- 446 465 473-475 484 487-489 492 498-499 502 504- 507 510 518 522 524 532 541 543 552 554-555 587- 588 595 602 610 631-632 643 651- 654 662 668-669 672-673
fetal brain	Clontech	FBR001	8 24 54 56 59 69 88 229 384 428 440-441 541 628 671
fetal brain	Clontech	FBR004	20 53 160-161 170 293 385 461 530 605 620-621 654 660
fetal brain	Clontech	FBR006	7-8 10 15 18-19 24-26 29 33 46 53 56 59 62-64 66 68

125
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			70 73 79 84 87 131 140 147-148 155 163 165 170 179- 180 189-190 208 211 218-219 229- 230 232-234 236 245 249 259-262 267 284-287 293 298 305 308 313- 314 316-319 322 324 337-338 343- 346 350-351 354 359-362 376 380- 381 384 387-398 403-404 423-424 428 431 435 438 440-441 445-447 451 462 473-475 484 492 498-501 504-507 509 512 516 518-519 521- 522 529-530 532 541 543 550 554 558 566 568-570 576 591 597-599 603 605 607-609 623-625 627-632 640 643 652-653 662-663 665 667 671-673 675 682
fetal brain	Clontech	FBRs03	17 371
fetal brain	Invitrogen	FBT002	7 10 29 43 47-49 52 60 64-65 67-68 79 83 86 92 94 131 139-140 168-169 180 202-203 205 218-219 230 242- 243 259-262 289 296 298 302 305 307 319 329-330 332-333 364 380 390 392-393 451 473-474 484 492 499-500 518 520 537 553 607 619 643 654
fetal heart	Invitrogen	FHR001	8 14-15 20 24-26 34 37 39 46 53 56- 57 60 63 70 75 80- 82 96-98 101 106 120 127 131 134 153 161 168-169 171 180 202-203 216 229 236 266- 267 289-290 303 305 308 314 316 325 344-345 356 358-359 363 366

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			371 384 392-393 395-398 400 402 419 422 431 434 436 438 451 453 461-462 478 484 500 504-506 518 522 525-526 530 535 537 539 541 550 570-573 586- 588 590-591 597 601 605 610 613- 614 626 630-632 640 643 652 669 672-673 675 682
fetal kidney	Clontech	FKD001	26 62 96 106 115 150 153 217-219 259-262 289 308 323-324 350 371 428 435 507 522 537 643
fetal kidney	Clontech	FKD002	46 54 64 68 85 107-108 126 131 155 158-159 163- 164 167-169 188 224-226 229 232- 234 236 245 282 284-285 289-290 293 298 340-341 343 350 370 417- 418 431 436 438 461 484 499-500 516 518 532 567- 568 572-574 589 596-599 613 624- 626 628 640 671- 673
fetal kidney	Invitrogen	FKD007	227
fetal lung	Clontech	FLG001	25 40 56 75 93 106 112-114 131 229 316 428 436 484 499 572-573 623
fetal lung	Invitrogen	FLG003	5 7 10 16 22 25-26 44 47-50 57 75 79 102 106 148 157 175-176 189 191 256 259-262 314 356 359 371 384 399-400 423-424 428 430 451 488- 490 500 504-507 518 529-530 534 539 550 556 620- 621
fetal lung	Clontech	FLG004	305
fetal liver-spleen	Columbia University	FLS001	1-5 7-8 10 12 14- 17 19-20 24-27 29- 54 56-57 62-64 68 71 75 80-83 85 87-

127
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			97 99-100 104-107 109-110 131 137 141-142 150-153 155 168-169 177- 180 183-184 188 198 200 202-203 205 208 212 215- 220 222 229 245 251-252 256-262 264 267 271-273 275-279 289-290 296 298 302 306 308 314 316-318 320 324-325 331- 333 337-338 349- 352 359 364 366- 368 371 377 383 386-387 390 392- 393 400-401 403- 404 420-421 423- 424 428 434-435 438 440-441 445- 446 451 455-457 459-462 475 479- 481 484 487 491- 492 498-507 510- 511 516 518 521- 524 526 530 533 536-538 541 543 550 554-556 558 588 593 595-598 601-602 605 607 610 613 620-621 623-624 629 634 641-643 651-652 667-668 671-673 675 681
fetal liver-spleen	Columbia University	FLS002	2-5 7-8 10 12 14- 17 19 24 26-27 34 36 38 40-42 44 47- 49 52-54 56-57 62 64 66 68 71 75-76 80-83 85-86 88-89 91-93 96 98-100 106-108 110 112- 113 115-117 128- 131 135 137 139- 142 150 153 157- 159 163 171-174 179 183-184 186 188-189 192 198 200 202-203 206 208 212 216 218- 220 229-230 236- 241 245 249 252 256-262 275-279 290 294-295 298 302 305-306 308

128

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			312-314 316-320 324-325 327 335 337-338 342 349- 350 357-360 366- 368 376-379 387 390 400 419-420 426 428 434 436 438 440-441 444 456-457 459-462 478 481-482 486- 492 498-499 501 504-506 508-509 516 518 521-522 527 530 534 536- 537 543 554-555 564 581 587-588 595 597-598 601 605 610 613 620- 621 623-625 627 629 631-632 634 641-643 651-652 662 666-668 671- 673 675 683
fetal liver- spleen	Columbia University	FLS003	2-5 14 18 20 24 26 44 62 64 68 80-83 88 93 99-100 106 137 153 157 163 183 197 222 229 236 245 256 275- 278 289 298 306 315-318 331 337- 338 346 350 359 366 371 419-420 428 436 438 491- 492 502 507 518 521-522 530 538 543 555-556 593 623-624 652 667 672-673 679
fetal liver	Invitrogen	FLV001	5 10 24 46 52 64 67-68 157 168-169 180 202-203 211 216 218-219 222 237-241 256 259- 262 272 275-278 317-318 321 324 332-333 342 347 351 371 401 421 428 434 451 488- 490 498 593 623 643 679
fetal liver	Clontech	FLV002	10 24 140 153 170 230 249 256 275- 278 284-285 325 358 366 392-393 500 518 538 576- 577 613 623 641- 642 666

129
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
fetal liver	Clontech	FLV004	5 13-14 18 20 24 35 46-50 56 63-64 68 75 100 102 106 108 116-118 137 140 144 147-148 170-172 218-219 236 256 259-262 275-278 318 323 325 329-330 340- 341 356-357 371 390 428 431 436 438 440-441 453 461-462 498-499 518 530 537-538 543 587-588 623 629 632 638-639 643 651-652 662 666 671-673
fetal muscle	Invitrogen	FMS001	5 16 24-26 64 93 139 144 168-169 171 175-176 181 202-203 212 218- 219 256 289-290 296 298 317-318 349 356 364 371 377 380 392-393 402-404 427 444 518 523 564 586 623 661-662
fetal muscle	Invitrogen	FMS002	6 15-16 21 26 29 37 41 52 57 75 87 96 101-102 106 116-118 131 158- 159 167-169 171 180 189 256-258 272 289-290 293 298 306 308 316 325 332-333 343 351 353 356 380 382 388 400 402 411 416 419 428- 429 431 453 499 516 522 525 530 532 541 543 550 563 565-568 572- 573 584 586 603 613 623 643 662- 663
fetal skin	Invitrogen	FSK001	5 7-8 10 14-17 20 23 25-26 29 36-37 39 41 46 51 53 68- 70 80-82 84 86 90 92-93 96 111 127- 130 132-133 141- 142 147-148 151- 152 158-161 163- 165 173-174 202- 203 205-207 218-

130

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			220 224-226 229- 230 245 254 256- 262 289-290 296 298 302 305 308- 309 315-316 319 324-325 327 358- 359 364 369 371 388 392-393 400 405-413 417-418 436 438 440-441 451 458-465 467- 472 476 487 492 499 518-520 525 530-532 547-549 558 564 571 580 583 591 607 610 617-618 620-621 623 627 643 652 659-663 671-673 680 682
fetal skin	Invitrogen	FSK002	5 10 16 18 20 23- 26 36-37 39 41 46 52-53 56 61-65 68 70 80-83 87 94 96 100 130-131 148 158-159 162-164 168-169 182 188 193 201 220 224- 226 229 235-241 245 249 254 257- 262 289-290 293 298 302 316 318 325 331-333 335 340-341 350 359 361 363-364 371 390 392-398 400 403-404 408-409 411 417-418 422 428 431 436 440- 441 451 453 462 464-465 467 471 476 478 484 499 502 504-506 512 516 518 521-522 530 532 541 543 547-549 556 564- 565 568 587 589- 591 593-594 597- 598 613-614 616 624-625 629 631- 632 637 640 643 652 662 667 669 671-673 681-682
fetal spleen	BioChain	FSP001	26 87 371 461 667
umbilical cord	BioChain	FUC001	5 18 20 26 40 47- 49 70 72 83 86-87 93 96 106 110-111 116-117 124 126-

131
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			127 134 144 152- 153 155 157-159 161 165 171 182 206 218-219 224- 226 229 243 247 249 256 259-262 289 296 298 303 305-306 308 314 316 325 332-333 337-338 344-345 349 352 359 364 371 394-398 400 417-421 427 431 436 438 453 473- 474 477 479 499- 500 507 512 522 525 535 537 565 593 595 613 620- 621 623-624 637 643 653-654 660- 661 668-669 682
fetal brain	GIBCO	HFB001	5 10 18-21 27 34 38-40 47-49 52 56- 60 62 64 66-70 72- 76 80 83 86 92-93 134 139 141-142 149-150 155 170 172 179-180 185- 186 188 202-203 205 207 209-212 216 229-230 256 286-287 289 294- 296 298 314 319- 320 323 325 337- 338 346 350 357 367 371 376 381 384 420 436 438 444 447 454 459- 462 475 484 487 492-493 499-500 507 518-519 522 529-530 532 534 541 543 563 570- 571 580 597-598 601 607 616 619- 621 623-624 643 653-654 662 664 668 671-673 675 677-678 682
macrophage	Invitrogen	HMP001	18 26 43 64 118 144 179 211 245 329-330 347 371 427 435 461 502 530 537 620-621 635 638-639
infant brain	Columbia University	IB2002	7 14 16-17 21 23 25-26 29 40 47-50 56-57 59-60 64 67-

132
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			68 70 73-74 79 83 88 91-92 94 98 103 115 127 137 139 150-152 156 158- 159 161-163 173- 174 182 186 188- 189 197 202-203 205-215 230 245 259-262 264 268 280 285 289 296 298 305 307-308 313-316 319 322- 324 326 334 346- 347 349-351 359 363-364 367 371 376-377 390 420 431 436 438 444 447-449 451 453 461-462 479 487 492 498-501 504- 506 516 519 522 529-530 537 541 543 545 556 564 572-573 588 592- 593 597-598 600 604-605 607 610 619 622 624 627- 628 643 652-654 660 663 674-675 682
infant brain	Columbia University	IB2003	7 10 16 19-20 25 29 35 43 46-50 56- 57 59-60 64 68 70 79-82 87 92 106 139 150 158-159 162-163 165 173- 174 181 186 189 202-203 205 210- 214 229-230 245 256 259-263 289- 290 298 305 307- 308 314-315 319 322 328 334 337- 338 347 349 351 359 364 371 380 385 428 436 438 444 447 449 451 462 475 484 487 492-493 498-502 519 522 529-530 532 537 540 550 556 593 602-605 607 616 622 627 631-632 643 652- 654 663 672-673 682
infant brain	Columbia University	IBM002	47-50 84 151-152 157 188-189 209

133
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			289 390 423-424 453 628
infant brain	Columbia University	IBS001	10 16 29 46-50 56 58 67 78 80-82 156 163 186 259-262 285 305 315 334 349 452 488-489 522 532 540
lung, fibroblast	Stratagene	LFB001	5 7 16 19 40 54 56 61-62 68 83 93 106 116 121 137 172 191 198 205 223 256 289 325 329 349 371 400 438 484 501-502 507 518 522 525 532 541 610 631-632 643 651 669
lung tumor	Invitrogen	LGT002	5-7 10 15-16 18-19 26 29 34-36 38 40- 41 46-50 52 56 59 64 68 75 86 89 91- 96 103-106 112-114 116-117 120 128- 130 135 141-142 144 147-148 150 154-155 157-159 162-164 172-174 179-180 190-192 198 202-203 208 215 220-221 223 229 236 249 255- 258 263 271 275- 278 284-285 291- 292 296 302 309 314 316 319 323 327 331 342 349- 351 353 358 364 368-369 371 390 392-393 399-400 420-421 427 431 436 438 444 453- 454 459-462 465 470 484 486 488- 492 499-500 502 507 511 518 522 525-526 530 537 539 543 550 580 597-599 605 623- 625 627 637 643 652 661-662 665- 666
lymphocytes	ATCC	LPC001	13 16 18 20 27 43 47-49 54 62-64 66- 68 80 87 90 96 98 115 118 120 131 144 163 202-203 211 252 256 259-

134

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			262 265 290 296 308 324-325 347 350 358 371 377 384 400 420 428 436 462 467 470 483-487 499-502 504-507 509 518 522 525 530 543 545 550 588 600 605 607 624-625 633 635 643 645 654 669 672-673 675
leukocyte	GIBCO	LUC001	10 16 18 24 34 38- 40 43-44 47-50 52 54-57 62-64 66 68 78 80-82 86-89 93- 94 98 106 109 111- 120 131 134 137 139 144 150-152 154 163 165 177 179 186 189 198 202-203 208 211 218-219 221 229 236 247 249 252 256 259-264 270 275-278 289-290 298 302 305 308 315 317-318 323 325 328 337-338 342 347 350 358 364 368 371 390 392-393 421 427- 428 430 433-435 437-438 440-441 444 451-452 454 461 475 484-487 491 493 498-500 502 504-507 509 518-519 522 525- 526 530 535 541 543 550 555 586- 588 597-598 605 607 610 620-621 624 627 631-633 638-639 643 652 654 668-669 672- 673 675-676
leukocyte	Clontech	LUC003	20 47-49 52 56 100 112-114 198-199 314 337-338 348 371 438 484 502 530 537 602 633 643
melanoma from- cell-line-ATCC- #CRL-1424	Clontech	MEL004	14 25 34 47-49 56 64 66 83 92 106 111 131 134 137 139 150 162 173-

135
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			174 189 192 210 229 249 259-262 290 321 337-338 350 364 371 392- 393 438 440-441 444 475 493 499 507 554 587 643 651 667 669 671
mammary gland	Invitrogen	MMG001	5 7 10 16-17 19 25 46-53 56 64 68 70 79-82 85-86 89 92- 95 98-100 106 121 127 137 139-142 144 150-152 158- 159 161-164 180 189 192-193 198 202-203 205-206 216 218-220 230 245 249 252 259- 263 267 270-272 275-278 289-290 298 302 305 308 313 315 319 324 329-330 336 346 349 351 355-356 359 364 368 370- 371 377 384 390 392-393 421 425 427-428 436 444 451 455-460 462 465 473-474 487 492 499 502 507 516 518 524-526 529-530 533-534 539 543 583 590 592 602 605 613 623 627 631-632 643 646 660 677- 678 682
induced neuron-cells	Stratagene	NTD001	17 20 23 68 79 89 153 155 181-182 212 218-219 235 298 346 352 358 376 438 478 484 488-489 492-493 499-501 541 570 619 627 643 662 672-673
retinoic acid-induced-neuronal-cells	Stratagene	NTR001	7 23 56 68 70 131 186 189 213-214 290 293 342 461 499 504-506 530 601-602 607 682
neuronal cells	Stratagene	NTU001	7 29 42 68 70 84- 85 92 131 140 147- 148 202-203 259- 262 305 316 319 336 371 395-398

136
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			461 493 499 502 537 550 553 592 652 672-673 682
pituitary gland	Clontech	PIT004	2-4 47-49 56 68 72 93 137-138 141-142 150 154 158-159 177 182 192 221 229 272-273 290 298 308 316 325 329-331 342 346 356 360 436 459- 460 462 473-474 484 504-507 524 532 534 541 543 564 623 631-632 635 643 662
placenta	Clontech	PLA003	1-5 7 12 26 37 41 53 64 75 85 87 96 106-107 112-114 131 151-152 157 223 236 256-262 303 306 316 335 350-351 359 371 400 428 431 435 438 445-446 462 499 502 516 520 530 532 537 543 550 556 565 579 587 594-595 626 635 638-639
prostate	Clontech	PRT001	20 25 56 173-174 205 250 256 280 284-285 299 302 309 320 323-324 331 342 349 362 367 384 386 392 400 415 438 484 498 507 524 532- 534 590 620-621 623 631-632 654 677-678 680
rectum	Invitrogen	REC001	7 10 20 47-50 52 85-87 89 109-110 126 128-130 157 163 170 173-174 177 205 220 229 256 259-262 289 319 324 327 340- 341 347 364 368 371 377 415-416 423-424 427 436 465 504-506 581- 582 602 610 679
salivary gland	Clontech	SAL001	5 10 22 25 43 52 63-64 67 89 95 97 99 137 140 161 165 167-169 180 205 229 252 256 290

137
TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			323 351 368 371 430 436 438 487 502 507 516 525 564 580 613 617- 618 631-632
saliva gland	Clontech	SALs03	20
skin fibroblast	ATCC	SFB001	208
skin fibroblast	ATCC	SFB002	208
small intestine	Clontech	SIN001	5 7-8 10 15 24 26 37-38 47-49 51-54 56-57 59 64 67-68 72 75 88 93 96-97 100 106 108 111 116-117 121 128- 131 137 140 153 158-159 177 189 191 202-203 206 215 229 253 255- 256 259-262 264- 265 272 280 296 300-301 308-309 316-318 325 327 332-333 335 337- 338 344-345 347 352 359 368 371 386 390 392-393 423 431 435 438 444 462 479 484 492 507 509 522 525-526 532 534 550 572-573 581 593 605 620-621 623 628 632 643 650 652-654 672- 673
skeletal muscle	Clontech	SKM001	5 62 101 104 134 165 254 272 289 300-301 308 316 323 356 377 402 428 431 438 444 451 462 541 543 550 572-573 586
skeletal muscle	Clontech	SKM002	208 507
spinal cord	Clontech	SPC001	13 15 26-27 33-34 38-40 46-50 52-53 56 68 80-82 87 89 92-95 131 150 155 163 175-176 180 186 197 199 202- 203 205 211 213- 214 229 231 235 254 263 289 307 311 314-316 323- 324 329 340-342 348-349 352 359 364 371 384 400 438 451 484 493 500 507 509 511

138

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			516 522 525 530 532 537 562-563 567-568 580 595 597 603 607 610 612-613 616 620- 622 627 643 653 672-673 675 677- 678
adult spleen	Clontech	SPLc01	7 9 13 17 26 37 43 64 75 106 112-114 118 131 163 212 216 218-219 256 259-262 308 314 329-330 349 368 390 392-393 422- 424 427 431 435- 436 451 453 484 500-501 509 525 530 532 535-536 541 592 600 610 613 623 628 631- 632 635 645 654 663 668 672-673 679
bone marrow	null	STM001	7 43 162 252 256 305 371 427 438 530 607 651 658
stomach	Clontech	STO001	67 93 95 135 230 259-262 284-285 289 302-303 308 320 323 390 392- 393 420 428 436 484 507 524-525 530 536 587 631- 632 637
thalamus	Clontech	THA002	10 18 24 33 47-50 54 58 60 68 90 92- 93 98 100 102 160- 161 180 205 208 229-230 242 259- 262 272 296 302- 305 325 331 342 359 384 386 390 425 511 532 543 572-573 587 602 608-610 612 616 620-621 631-632 660
thymus	Clontech	THM001	5 12 39-40 43 47- 50 54 56 66 68 70 79 87-88 93 106- 107 131 135 144 162 173-174 177 192 198 205 211 218-219 229 256 281 289-290 293 306 308 314 317- 318 321 323 325

139

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			331-333 347 349- 352 368 371 384 389 420 425 438 440-441 484 487 493 498-499 502 509 530 532 541 554-555 558 597- 599 610 613 616 620-621 624 643 671-673 682
thymus	Clontech	THMc02	5 8 10 12 25 32 34 37 39 43 45-46 48- 50 53 55-56 61 63 65-67 70 83 85 87- 88 94 106-107 112- 114 116-118 120 131 135 140-142 144 150-152 158- 159 163-165 179 189 208 229 232- 234 256 259-262 273 289-290 302 305 316-318 324- 325 335 349 361 363-364 371 384 389 392-393 421- 424 437-441 443 445-446 451 459- 461 473-474 498 500 504-507 509 518 522 526 530 541 554 564 583 592 600 607 610 613 624-625 627 630-632 634 637 643-645 651 667 669 671-673 682
thyroid gland	Clontech	THR001	6 14-15 19 26 29 32 34 39-40 47-52 56 61-63 66-68 72 75 87 93 95 100 104-106 115 128- 131 137 141-142 154 157 162 165 168-169 175 177 182 189 191-193 202-203 211 217- 219 221 229 231- 234 249 254 256 282 289-290 298 302 306-308 314- 316 323-324 327 329-330 342 350 353-358 368 371 377 380 383-384 400 423-424 426 431 436-438 440- 441 446 451 459-

140

TABLE 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			461 475 478 484 487-489 491-492 499-500 502-506 509 518-519 521- 522 530 532-533 541 543 567-568 586 588 597-600 605 607 610 617- 618 620-621 624- 626 631-632 635 643 651 654 662 668 671-672 680
trachea	Clontech	TRC001	7 22 38 40 56 68 83 94 229 259-262 289 296 298 360 371-375 438 484 499 511 521 541 571-573 588 613 624 627
uterus	Clontech	UTR001	17 36 70 76 103 106 109 112-114 131 150 157 179- 180 189 290 296 308 314 320 329- 330 356 364 366 368 390 395-398 415 438 447 507 509 519 525 529 532 564 620-621 631-632 662 668- 669 682

*The 16 tissue/mRNAs and their vendor sources are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) Normal adult kidney mRNA (Invitrogen), 3) Normal fetal brain mRNA (Invitrogen), 4) Normal adult liver mRNA (Invitrogen), 5) Normal fetal kidney mRNA (Invitrogen), 6) Normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) Human bone marrow mRNA (Clontech), 10) Human leukemia lymphoblastic mRNA (Clontech), 11) Human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human so\spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
685	gi183150	Homo sapiens	chorionic somatomammotropin CS-5	320	100
685	gi181127	Homo sapiens	chorionic somatomammotropin precursor	275	96
685	gi183153	Homo sapiens	chorionic somatomammotropin CS-2	275	96
686	gi183178	Homo sapiens	hGH-V2	1033	78
686	gi183153	Homo sapiens	chorionic somatomammotropin CS-2	710	87
686	gi387024	Homo sapiens	placental lactogen hormone precursor	710	87
688	gi183178	Homo sapiens	hGH-V2	1051	79
688	gi181121	Homo sapiens	chorionic somatomammotropin	788	95
688	gi183151	Homo sapiens	chorionic somatomammotropin CS-1	788	95
689	gi12653501	Homo sapiens	Similar to serine (or cysteine) proteinase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor). member 1	1242	99
689	gi15217079	Homo sapiens	pigment epithelium-derived factor	1242	99
689	gi189778	Homo sapiens	pigment epithelial-differentiating factor	1242	99
690	gi17128288	synthetic construct	Primer 1	1150	99
690	gi20269957	Sus scrofa	phospholipase C delta 4	1033	88
690	gi21307610	Mus musculus	phospholipase C delta 4	909	77
691	gi17864023	Homo sapiens	KCCR13L	3524	100
691	gi21483462	Drosophila melanogaster	LD44686p	533	36
691	gi21741717	Oryza sativa	oj991113_30.22	127	29
692	gi17428818	Ralstonia solanacearum	GALA PROTEIN 3	117	32
692	gi21536497	Arabidopsis thaliana	F-box protein family, AtFBL4	115	30
692	gi12581504	Trypanosoma brucei	GU1	115	33
693	gi437662	Oryctolagus cuniculus	interleukin-8 receptor subtype B	194	61
693	gi186378	Homo sapiens	interleukin 8 receptor B	178	57
693	gi1109691	Homo sapiens	interleukin-8 receptor type B	178	57
694	gi3335098	Homo sapiens	CD39L2	2520	100
694	gi11230487	Rattus norvegicus	NTPDase6	2065	86
694	gi5139519	Mus musculus	nucleoside diphosphatase (ER-UDPase)	1008	53
695	gi21928620	Homo sapiens	seven transmembrane helix receptor	1858	100
695	gi16566319	Homo sapiens	G protein-coupled receptor	1843	99
695	gi6644328	Rattus norvegicus	orphan G protein-coupled receptor GPR26	822	50
696	gi7110216	Homo sapiens	C-type lectin-like receptor-1	851	99
696	gi7109731	Homo sapiens	C-type lectin-like receptor-2	256	31
696	gi20381202	Mus musculus	Similar to C-type (calcium dependent, carbohydrate recognition domain) lectin, superfamily member 12	196	27
697	gi22449809	Chaoborus trivitattus	cytochrome oxidase I	50	44
697	gi2351328	Newcastle disease virus	fusion protein	59	44
697	gi21311450	Galleria mellonella	antifungal peptide gallerimycin	55	33
698	gi18089247	Homo sapiens	Similar to ectonucleoside triphosphate diphosphohydrolase 5	2104	100

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
698	gi3335102	Homo sapiens	CD39L4	2104	100
698	gi15076827	Homo sapiens	Pcph proto-oncogene protein	2090	99
699	gi151242	Pseudomonas aeruginosa	heat shock protein	79	38
699	gi9950616	Pseudomonas aeruginosa	GroES protein	79	38
699	gi2564287	Pseudomonas stutzeri	Hsp10 protein	79	44
701	gi20521055	Homo sapiens	Start codon is not identified	724	32
701	gi17225457	Homo sapiens	autism-related protein 1	676	32
701	gi15145797	Sus scrofa	basic proline-rich protein	156	27
702	gi20810589	Homo sapiens	similar to arsenite inducible RNA associated protein	833	99
702	gi9651711	Mus musculus	arsenite inducible RNA associated protein	687	80
702	gi17390981	Homo sapiens	Similar to RIKEN cDNA 1110060O18 gene	535	59
703	gi6624130	Rattus norvegicus	similar to 45 kDa secretory protein ;	2150	100
703	gi13241652	Rattus norvegicus	supernatant protein factor	2040	93
703	gi19548982	Bos taurus	tocopherol-associated protein	1930	90
704	gi13177766	Homo sapiens	Similar to presenilins associated rhomboid-like protein	1761	99
704	gi15559382	Homo sapiens	presenilins associated rhomboid-like protein	1094	98
704	gi7959883	Homo sapiens	PRO2207	671	82
705	gi1864091	Rattus norvegicus	PSD-95/SAP90-associated protein-3	5005	95
705	gi2454510	Homo sapiens	PSD-95/SAP90-associated protein-2	1338	55
705	gi6979173	Homo sapiens	discs, large (Drosophila) homolog-associated protein 2	1011	45
706	gi11877274	Homo sapiens	dJ726C3.2 (novel protein)	2260	99
706	gi21667210	Homo sapiens	bactericidal/permeability-increasing protein-like 1	2260	99
706	gi20387087	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein) like-2	349	26
707	gi7291716	Drosophila melanogaster	CG11388-PA	648	39
707	gi16768190	Drosophila melanogaster	GH22974p	647	39
707	gi3954938	Homo sapiens	acetylglucosaminyltransferase-like protein	171	23
708	gi14334082	Mus musculus	thymus LIM protein TLP-A	479	87
708	gi14334084	Mus musculus	thymus LIM protein TLP-B	397	79
708	gi487284	Rattus norvegicus	CRP2 (cysteine-rich protein 2)	367	75
710	gi556299	Mus musculus	alpha-2 type IV collagen	8129	83
710	gi30076	Homo sapiens	alpha-2 chain precursor (AA -25 to 1018) (3416 is 2nd base in codon)	5916	100
710	gi15991848	Homo sapiens	A type IV collagen	4239	51
711	gi7861733	Homo sapiens	low density lipoprotein receptor related protein-deleted in tumor	2583 1	99
711	gi8926243	Mus musculus	low density lipoprotein receptor related protein LRP1B/LRP-DIT	2409 6	91

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
711	gi438007	Gallus gallus	alpha-2-macroglobulin receptor	14197	63
712	gi17298315	Homo sapiens	candidate tumor suppressor protein	848	100
712	gi7861733	Homo sapiens	low density lipoprotein receptor related protein-deleted in tumor	848	100
712	gi8926243	Mus musculus	low density lipoprotein receptor related protein LRP1B/LRP-DIT	731	83
713	gi16877754	Homo sapiens	Similar to RIKEN cDNA 4930434H03 gene	574	56
713	gi20071811	Mus musculus	Similar to RIKEN cDNA 4930434H03 gene	493	60
713	gi1340174	Homo sapiens	type III procollagen (aa 892-1023)	97	40
714	gi157409	Drosophila melanogaster	fat protein	1802	31
714	gi4887715	Drosophila melanogaster	adherin	1500	36
714	gi1107687	Homo sapiens	homologue of Drosophila Fat protein	1514	30
715	gi157409	Drosophila melanogaster	fat protein	1808	31
715	gi4887715	Drosophila melanogaster	adherin	1500	36
715	gi1107687	Homo sapiens	homologue of Drosophila Fat protein	1514	30
716	gi17865311	Homo sapiens	dipeptidyl peptidase-like protein 9	2562	99
716	gi3513303	Homo sapiens	R26984_1	2700	98
716	gi11095188	Homo sapiens	dipeptidyl peptidase 8	1397	53
717	gi2689444	Homo sapiens	ZNF134	1160	54
717	gi21314977	Homo sapiens	Similar to zinc finger protein 17 (HPF3, KOX 10)	1038	51
717	gi13543419	Homo sapiens	Similar to zinc finger protein 304	1000	51
718	gi7582294	Homo sapiens	BM-011	881	100
718	gi13937769	Homo sapiens	Similar to RIKEN cDNA 1200013F24 gene	781	98
718	gi178997	Homo sapiens	arginine-rich nuclear protein	224	38
719	gi1620870	Ciona intestinalis	myoplasmin-C1	412	28
719	gi7416980	Argopecten irradians	myosin heavy chain catch (smooth) muscle specific isoform	279	23
719	gi7416982	Argopecten irradians	myosin heavy chain cardiac muscle specific isoform 1	279	23
720	gi13872813	Homo sapiens	fibulin-6	13764	100
720	gi14575679	Homo sapiens	hemicentin	13720	99
720	gi3328186	Caenorhabditis elegans	hemicentin precursor	1695	30
721	gi3822553	Gallus gallus	nuclear calmodulin-binding protein	1492	64
721	gi3329496	Mus musculus	heterogenous nuclear ribonucleoprotein U	1501	45
721	gi624918	Rattus norvegicus	SP120	1498	45
722	gi17223626	Homo sapiens	ATP-binding cassette A10	7966	99
722	gi17223624	Homo sapiens	ATP-binding cassette A9	5160	61
722	gi17223622	Homo sapiens	ATP-binding cassette A6	5108	61
723	gi13374079	Homo sapiens	TAFII140 protein	3677	99
723	gi13374178	Mus musculus	TAFII140 protein	3202	84

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
723	gi205686	Rattus norvegicus	heavy neurofilament subunit	335	26
724	gi17429038	Ralstonia solanacearum	PROBABLE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE PROTEIN	661	61
724	gi9948609	Pseudomonas aeruginosa	probable acyl-CoA dehydrogenase	619	62
724	gi13421911	Caulobacter crescentus CB15	acyl-CoA dehydrogenase family protein	559	59
725	gi6752658	Homo sapiens	epidermal growth factor repeat containing protein	3055	99
725	gi16040981	Mus musculus	POEM	884	51
725	gi15430246	Mus musculus	nephronectin short isoform	884	51
726	gi6531661	Caenorhabditis elegans	LIN-41A	844	50
726	gi6531663	Caenorhabditis elegans	LIN-41B	844	50
726	gi12407367	Homo sapiens	tripartite motif protein TRIM2	769	30
727	gi1504026	Homo sapiens	similar to C.elegans protein (Z37093)	5833	99
727	gi2896796	Homo sapiens	D1013901	5115	99
727	gi2522322	Homo sapiens	PTPL1-associated RhoGAP	1497	36
728	gi13274120	Homo sapiens	dJ55C23.5.1 (vanin 3, isoform 1)	1467	99
728	gi7160973	Homo sapiens	VNN3 protein	1213	96
728	gi6102996	Mus musculus	Vanin-3	1018	79
729	gi9581879	Homo sapiens	disintegrin metalloproteinase with thrombospondin repeats	5723	99
729	gi19171176	Homo sapiens	metalloprotease disintegrin 15 with thrombospondin domains	1669	50
729	gi11095299	Rattus norvegicus	ADAMTS-1	1772	40
730	gi21063967	Drosophila melanogaster	AT05453p	396	32
730	gi5911409	Drosophila melanogaster	fuzzy	396	32
730	gi2564657	Drosophila melanogaster	Fuzzy	396	32
731	gi15217171	Homo sapiens	CD81 partner 3	2302	100
731	gi15488017	Homo sapiens	EWI2	2302	100
731	gi15593237	Mus musculus	immunoglobulin superfamily receptor PGRL	2186	92
732	gi15217171	Homo sapiens	CD81 partner 3	3200	100
732	gi15488017	Homo sapiens	EWI2	3200	100
732	gi15593237	Mus musculus	immunoglobulin superfamily receptor PGRL	2867	88
733	gi15217171	Homo sapiens	CD81 partner 3	1303	96
733	gi15488017	Homo sapiens	EWI2	1303	96
733	gi22266726	Homo sapiens	LIR-D1 precursor	1303	96
734	gi21748480	Homo sapiens	FLJ00271 protein	605	100
734	gi22266726	Homo sapiens	LIR-D1 precursor	514	79
734	gi15217171	Homo sapiens	CD81 partner 3	514	79
735	gi2196872	Homo sapiens	Lsc homologue	203	30
735	gi1389756	Mus musculus	Lsc	199	31
735	gi11276027	Rattus	LSC	199	31

145

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		norvegicus			
736	gi14336728	Homo sapiens	possible integral membrane	331	32
736	gi18043242	Mus musculus	RIKEN cDNA 2400010G15 gene	331	31
736	gi8895014	Hepatitis B virus	HBsAg	68	48
737	gi20071204	Mus musculus	Similar to paraspeckle protein 1	185	28
737	gi18104577	Homo sapiens	paraspeckle protein 1 alpha isoform	175	27
737	gi13528666	Homo sapiens	Similar to splicing factor proline/glutamine rich (polypyrimidine tract-binding protein-associated)	179	31
738	gi12002000	Homo sapiens	My029 protein	415	100
738	gi348140	Human T-lymphotropic virus 2	rex	68	39
738	gi404041	Human T-lymphotropic virus 2	rex protein	68	39
739	gi4680090	Human immunodeficiency virus type 1	envelope glycoprotein	89	31
740	gi21627272	Drosophila melanogaster	CG12765-PA	166	38
740	gi19528077	Drosophila melanogaster	AT24025p	166	38
740	gi1066820	Murray Valley encephalitis virus	nonstructural protein	66	28
741	gi9916	Plasmodium falciparum	liver stage antigen	468	26
741	gi1747	Oryctolagus cuniculus	trichohyalin	414	24
741	gi295941	Ovis aries	trichohyalin	395	24
742	gi9845485	Homo sapiens	protocadherin-9	6235	100
742	gi15054521	Homo sapiens	protocadherin-S	3390	58
742	gi13161060	Homo sapiens	protocadherin 11	3382	58
743	gi5688958	Homo sapiens	PMMLP	2405	100
743	gi21594625	Mus musculus	RIKEN cDNA 4931406N15 gene	2241	92
743	gi16797814	Drosophila melanogaster	phosphomannomutase 45A	1194	51
744	gi21734445	Rattus norvegicus	BMP/Retinoic acid-inducible neural-specific protein-2	3987	94
744	gi20988899	Mus musculus	similar to deleted in bladder cancer chromosome region candidate 1	2952	70
744	gi21734447	Rattus norvegicus	BMP/Retinoic acid-inducible neural-specific protein-3	2951	70
745	gi2739353	Homo sapiens	ZNF91L	2075	69
745	gi1017722	Homo sapiens	repressor transcriptional factor	2044	71
745	gi4559318	Homo sapiens	BC273239_1	2031	67
746	gi1017722	Homo sapiens	repressor transcriptional factor	2144	73
746	gi2739353	Homo sapiens	ZNF91L	2054	70
746	gi186774	Homo sapiens	zinc finger protein	2035	70
747	gi19683999	Homo sapiens	coated vesicle membrane protein	1010	99
747	gi1212965	Homo sapiens	transmembrane protein	1010	99
747	gi1213221	Rattus norvegicus	transmembrane protein	1006	98

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
748	gi1199524	Homo sapiens	acid phosphatase	2036	98
748	gi34263	Homo sapiens	acid phosphatase precursor protein	2036	98
748	gi13111975	Homo sapiens	acid phosphatase 2, lysosomal	2032	98
749	gi15625570	Homo sapiens	centaurin beta5	2970	83
749	gi4688902	Homo sapiens	centaurin beta2	1708	64
749	gi436228	Homo sapiens	Start codon is not identified	1387	70
750	gi10197642	Homo sapiens	MDS022	647	100
750	gi19683046	Dictyostelium discoideum	HYPOTHETICAL 21.8 KDA PROTEIN. 3/101	94	26
750	gi6841554	Homo sapiens	HSPC166	93	24
751	gi5630080	Homo sapiens	similar to HUB1; similar to BAA24380 (PID:g2789430)	696	48
751	gi2789430	Homo sapiens	repressor protein	702	39
751	gi18614026	Homo sapiens	zinc finger DNA binding protein p71	1004	41
752	gi12140290	Homo sapiens	bA12M19.2.1 (vacuolar protein sorting protein 16 (VPS16))	2885	92
752	gi11345382	Homo sapiens	vacuolar protein sorting protein 16	2885	92
752	gi19343731	Mus musculus	vacuolar protein sorting 16 (yeast homolog)	2803	89
753	gi20987877	Mus musculus	similar to Nogo receptor	905	58
753	gi9280025	Macaca fascicularis	Nogo receptor	808	49
753	gi15080005	Homo sapiens	nogo receptor	796	48
754	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2714	39
754	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	2708	39
754	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	2700	39
755	gi4929790	Homo sapiens	angiopoietin-related protein 3	1423	89
755	gi13159474	Homo sapiens	CG006-alt2	1416	88
755	gi5639997	Mus musculus	angiopoietin-related protein 3	1109	77
756	gi200057	Mus musculus	neuronal glycoprotein	4821	87
756	gi563133	Rattus norvegicus	BIG-1 protein	4778	87
756	gi1016012	Rattus norvegicus	neural cell adhesion protein BIG-2 precursor	3867	68
757	gi6273399	Homo sapiens	melanoma-associated antigen MG50	344	33
757	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	344	33
757	gi14495561	Homo sapiens	brain tumor associated protein LRRC4	324	27
758	gi6273399	Homo sapiens	melanoma-associated antigen MG50	344	33
758	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	344	33
758	gi14495561	Homo sapiens	brain tumor associated protein LRRC4	329	26
759	gi5525078	Rattus norvegicus	seven transmembrane receptor	5062	72
759	gi21929093	Homo sapiens	seven transmembrane helix receptor	1712	88
759	gi4164023	Bos taurus	latrophilin 2 splice variant baaaf	383	27
760	gi10440398	Homo sapiens	FLJ00032 protein	1261	57
760	gi11917507	Homo sapiens	HPF1 protein	1258	60
760	gi13752754	Homo sapiens	zinc finger 1111	1253	60
761	gi3628757	Homo sapiens	FIC1	1436	54
761	gi13097633	Homo sapiens	Similar to ATPase, Class I, type 8B, member 1	1221	60
761	gi20147219	Arabidopsis thaliana	Atlg59820/F23H11_14	1637	41

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
762	gi11527987	Gallus gallus	immunoglobulin-like receptor CHIR-A	97	30
762	gi432214	Human immunodeficiency virus type 1	envelope glycoprotein gp120	43	39
762	gi15026993	Homo sapiens	MUC5AC protein	64	38
763	gi11558486	Homo sapiens	B-cell lymphoma/leukaemia 11A short form	1314	99
763	gi7546791	Mus musculus	CTIP1 protein	1149	99
763	gi7650184	Mus musculus	ecotropic viral integration site 9 isoform C	1155	95
764	gi22085890	Rattus norvegicus	FHA-HIT	1426	82
764	gi21430028	Drosophila melanogaster	GM01362p	338	40
764	gi21166012	Dictyostelium discoideum	2410016G21RIK PROTEIN	279	26
765	gi22085890	Rattus norvegicus	FHA-HIT	214	88
765	gi5764101	Homo sapiens	polynucleotide kinase-3'-phosphatase	95	50
765	gi5712131	Homo sapiens	DEM1 protein	93	50
766	gi22085890	Rattus norvegicus	FHA-HIT	278	89
766	gi5764101	Homo sapiens	polynucleotide kinase-3'-phosphatase	109	46
766	gi5712131	Homo sapiens	DEM1 protein	107	46
768	gi15186770	Homo sapiens	lysyl oxidase-like protein	1818	96
768	gi14009597	Homo sapiens	lysyl oxidase-like 3 protein	1818	96
768	gi15030096	Mus musculus	Similar to lysyl oxidase-like 2	1715	92
769	gi3954938	Homo sapiens	acetylglucosaminyltransferase-like protein	2298	70
769	gi3954978	Mus musculus	acetylglucosaminyltransferase-like protein	2298	70
769	gi10834722	Homo sapiens	PP5656	892	91
770	gi7209723	Homo sapiens	WD-repeat like sequence	2476	99
770	gi8217485	Homo sapiens	dJ1092A11.3 (WD repeat domain)	2473	99
770	gi7209721	Mus musculus	DD57	2243	88
771	gi18676632	Homo sapiens	FLJ00215 protein	1943	99
771	gi18447198	Drosophila melanogaster	GH09355p	140	19
771	gi295671	Saccharomyces cerevisiae	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III	119	22
772	gi10799166	Homo sapiens	protein kinase Njmu-R1	1915	99
772	gi21104460	Homo sapiens	OK/SW-CL.19	549	100
772	gi14290030	Human immunodeficiency virus type 1	pol protein	68	30
773	gi4186023	Homo sapiens	CDS2 protein	2376	100
773	gi19344052	Homo sapiens	similar to PHOSPHATIDATE CYTIDYLYLTRANSFERASE 2 (CDP-DIGLYCERIDE SYNTHETASE 2) (CDP-DIGLYCERIDE PYROPHOSPHORYLASE 2) (CDP-DIACYLGLYCEROL SYNTHASE 2) (CDS 2) (CTP:PHOSPHATIDATE CYTIDYLYLTRANSFERASE 2) (CDP-	2376	100

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
			DAG SYNTHASE 2) (CDP-DG SYNTHETASE 2)...		
773	gi13277972	Mus musculus	Similar to CDP-diacylglycerol synthase (phosphatidate cytidyltransferase) 2	2289	96
774	gi17862928	Drosophila melanogaster	SD03549p	125	35
774	gi18077663	Mus musculus	cockayne syndrome group A	117	38
774	gi14091657	Mangifera indica	F6N15.8-like protein	107	29
776	gi18676664	Homo sapiens	FLJ00231 protein	1473	99
776	gi16303748	Homo sapiens	tweety-like protein 2	1053	41
776	gi16303750	Mus musculus	tweety homolog 2	987	39
777	gi8118032	Homo sapiens	orphan G-protein coupled receptor	939	98
777	gi16877193	Homo sapiens	G protein-coupled receptor, family C, group 5, member C	939	98
777	gi9588669	Homo sapiens	GPRC5C	939	98
778	gi20380605	Mus musculus	RIKEN cDNA 8430424D23 gene	836	91
778	gi16769562	Drosophila melanogaster	LD38910p	333	47
778	gi7302978	Drosophila melanogaster	CG8441-PA	333	47
779	gi16041781	Homo sapiens	Similar to RIKEN cDNA 0710001C05 gene	776	99
779	gi21430012	Drosophila melanogaster	GH27470p	333	53
779	gi15074454	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	239	43
780	gi13959018	Homo sapiens	endothelial cell-selective adhesion molecule	902	100
780	gi13991773	Mus musculus	endothelial cell-selective adhesion molecule	643	70
780	gi1814277	Homo sapiens	A33 antigen precursor	229	34
781	gi8164184	Homo sapiens	22kDa peroxisomal membrane protein-like	1013	100
781	gi15422171	Homo sapiens	22 kDa peroxisomal membrane protein 2	1013	100
781	gi297437	Rattus norvegicus	peroxisomal membrane protein	798	76
782	gi7621329	Streptococcus pyogenes	Sic1.245	214	39
782	gi7620883	Streptococcus pyogenes	Sic1.23	215	39
782	gi7620875	Streptococcus pyogenes	Sic1.19	215	39
783	gi62877	Gallus gallus	type VI collagen alpha-2 subunit preprotein	751	41
783	gi62882	Gallus gallus	type VI collagen subunit alpha2	751	41
783	gi211616	Gallus gallus	type VI collagen, alpha-2 subunit	747	45
784	gi17945608	Drosophila melanogaster	RE26969p	829	48
784	gi3877350	Caenorhabditis elegans	contains similarity to Pfam domain: PF01598 (Sterol desaturase), Score=307.6, E-value=4.7e-89, N=1	572	38
784	gi3877351	Caenorhabditis elegans	contains similarity to Pfam domain: PF01598 (Sterol desaturase), Score=303.0, E-value=1.1e-87, N=1	546	38

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
785	gi17066106	Homo sapiens	Novex-3 Titin Isoform	8832	99
785	gi21238628	Sparisoma viride	titin-like protein	519	62
785	gi21238630	Sparisoma aurofrenatum	titin-like protein	519	62
787	gi2230840	Ginkgo biloba	ndhB	54	54
787	gi2230828	Dioon edule	ndhB	52	50
787	gi9279991	Sequoia sempervirens	maturase	60	36
788	gi18676610	Homo sapiens	FLJ00204 protein	204	27
788	gi3002588	Mus musculus	Plenty of SH3s; POSH	206	25
788	gi1407665	Mus musculus	SH3P3	134	45
789	gi18676610	Homo sapiens	FLJ00204 protein	262	27
789	gi3002588	Mus musculus	Plenty of SH3s; POSH	220	25
789	gi1407665	Mus musculus	SH3P3	140	33
790	gi182483	Homo sapiens	prefibroblast collagenase inhibitor	531	88
790	gi490094	Homo sapiens	TIMP	531	88
790	gi189382	Homo sapiens	collagenase inhibitor	531	88
791	gi7110216	Homo sapiens	C-type lectin-like receptor-1	851	99
791	gi7109731	Homo sapiens	C-type lectin-like receptor-2	256	31
791	gi1902982	Bos taurus	lectin-like oxidized LDL receptor	303	31
792	gi5802604	Cavia porcellus	UDP glucuronosyltransferase UGT2A3	1783	73
792	gi19387963	Mus musculus	RIKEN cDNA 2010321J07 gene	1709	69
792	gi4753766	Homo sapiens	UDP glucuronosyltransferase	1598	67
793	gi3688090	Homo sapiens	R32611_2	786	91
793	gi6841228	Homo sapiens	HSPC289	638	78
793	gi21618688	Mus musculus	RIKEN cDNA 5830498C14 gene	445	52
794	gi9963861	Homo sapiens	Cyt19	1729	99
794	gi15488645	Mus musculus	methyltransferase Cyt19	1555	76
794	gi18150409	Rattus norvegicus	S-adenosylmethionine:arsenic (III) methyltransferase	1516	76
795	gi11877243	Homo sapiens	SSF1/P2Y11 chimeric protein	1957	95
795	gi21619996	Homo sapiens	peter pan homolog (Drosophila)	2080	99
795	gi14602631	Homo sapiens	peter pan (Drosophila) homolog	2080	99
796	gi20330550	Homo sapiens	NK inhibitory receptor precursor	799	98
796	gi20380183	Homo sapiens	similar to CMRF35 leukocyte immunoglobulin-like receptor	727	92
796	gi20381405	Homo sapiens	similar to CMRF35 leukocyte immunoglobulin-like receptor; CMRF35 antigen	423	57
797	gi20330550	Homo sapiens	NK inhibitory receptor precursor	799	98
797	gi20380183	Homo sapiens	similar to CMRF35 leukocyte immunoglobulin-like receptor	727	92
797	gi20381405	Homo sapiens	similar to CMRF35 leukocyte immunoglobulin-like receptor; CMRF35 antigen	423	57
798	gi20330550	Homo sapiens	NK inhibitory receptor precursor	1469	94
798	gi20380183	Homo sapiens	similar to CMRF35 leukocyte immunoglobulin-like receptor	690	84
798	gi20330544	Mus musculus	polymeric immunoglobulin receptor 3 precursor	416	52
799	gi18307481	Homo sapiens	phosphoinositide-binding proteins	2122	100
799	gi3930781	Homo sapiens	connector enhancer of KSR-like protein CNK1	346	34

150

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
799	gi4151807	Rattus norvegicus	membrane-associated guanylate kinase-interacting protein 2 Maguin-2	455	37
800	gi15929988	Homo sapiens	Similar to TLH29 protein precursor	417	89
800	gi11493982	Homo sapiens	TLH29 protein precursor	274	72
800	gi20147034	Mus musculus	interferon stimulated gene 12	235	68
801	gi15929988	Homo sapiens	Similar to TLH29 protein precursor, clone MGC:21991 IMAGE:4398045, mRNA, complete cds.	445	100
801	AAW54040	Homo sapiens	Human interferon-inducible protein, HIFI.	432	97
801	gi11493982	Homo sapiens	TLH29 protein precursor (TLH29) mRNA, complete cds.	303	70
802	gi12082725	Mus musculus	B cell phosphoinositide 3-kinase adaptor	3561	84
802	gi12082723	Gallus gallus	B cell phosphoinositide 3-kinase adaptor	2840	69
802	gi20987486	Homo sapiens	similar to B cell phosphoinositide 3-kinase adaptor	1830	97
803	gi7959809	Homo sapiens	PRO1082	545	100
803	gi7767407	Avian infectious bronchitis virus	5a protein	61	26
803	gi15073792	Sinorhizobium meliloti	PUTATIVE FOSMIDOMYCIN RESISTANCE ANTIBIOTIC RESISTANCE TRANSMEMBRANE PROTEIN	71	38
804	gi15384843	Homo sapiens	NTB-A receptor	1700	100
804	gi15384841	Homo sapiens	activating NK receptor	1687	99
804	gi9887089	Mus musculus	lymphocyte antigen 108 isoform 1	637	44
805	gi17979255	Arabidopsis thaliana	AT5g49550/K6M13_10	211	72
805	gi10177621	Arabidopsis thaliana	phytoene dehydrogenase-like	195	75
805	gi14023915	Mesorhizobium loti	phytoene dehydrogenase	182	62
806	gi14270364	Mus musculus	Epigen protein	386	71
806	gi755468	Xenopus laevis	transmembrane protein	120	36
806	gi7799191	Mus musculus	tomoregulin-1	125	52
807	gi14270364	Mus musculus	Epigen protein	386	71
807	gi755468	Xenopus laevis	transmembrane protein	120	36
807	gi7799191	Mus musculus	tomoregulin-1	125	52
808	gi14270364	Mus musculus	Epigen protein	386	71
808	gi755468	Xenopus laevis	transmembrane protein	120	36
808	gi7799191	Mus musculus	tomoregulin-1	125	52
809	gi3068592	Mus musculus	punc	201	41
809	gi22003417	Danio rerio	neogenin	193	40
809	gi1881477	Mus musculus	neogenin protein	167	33
810	gi15072404	Raja erinacea	organic solute transporter beta	92	41
810	gi143486	Bacillus subtilis	levansucrase	59	37
810	gi143484	Bacillus subtilis	levansucrase (sacB)	58	35
811	gi18650588	Homo sapiens	retinoic acid early transcript 1	1124	99
811	gi13128925	Homo sapiens	ULBP2 protein	1070	94
811	gi21961213	Homo sapiens	UL16 binding protein 2	1070	94
812	gi9280405	Homo sapiens	adlican	1372	46
812	gi3328186	Caenorhabditis elegans	hemicentin precursor	475	29

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
812	gi14575679	Homo sapiens	hemicentin	493	28
814	gi9280405	Homo sapiens	adlican	2438	35
814	gi14575679	Homo sapiens	hemicentin	688	25
814	gi3328186	Caenorhabditis elegans	hemicentin precursor	586	26
815	gi21619635	Homo sapiens	similar to Alu subfamily SQ sequence contamination warning entry	270	60
815	gi6650810	Homo sapiens	PRO1902	264	63
815	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	247	62
816	gi6707435	Homo sapiens	apolipoprotein A5	1864	100
816	gi12240284	Mus musculus	apolipoprotein A5	1310	72
816	gi6707431	Rattus norvegicus	apolipoprotein A5	1293	72
817	gi6707435	Homo sapiens	apolipoprotein A5	1864	100
817	gi12240284	Mus musculus	apolipoprotein A5	1310	72
817	gi6707431	Rattus norvegicus	apolipoprotein A5	1293	72
818	gi12751065	Homo sapiens	PNAS-25	360	81
818	gi1208732	Drosophila melanogaster	ovary2	276	33
818	gi21428518	Drosophila melanogaster	LD33046p	275	33
819	gi5771420	Homo sapiens	group IID secretory phospholipase A2	852	100
819	gi6453793	Homo sapiens	phospholipase A2	846	99
819	gi10862736	Homo sapiens	dJ169O23.3 (phospholipase A2 group IID)	846	99
820	gi6015448	Hylobates lar	dopamine receptor D4	79	35
820	gi5059331	Human papillomavirus type 83	major capsid protein	85	29
820	gi13278034	Mus musculus	Similar to selectin, platelet (p-selectin) ligand	83	35
821	gi12654883	Homo sapiens	rTS beta protein	2112	96
821	gi1150421	Homo sapiens	rTSbeta	2112	96
821	gi11094019	Homo sapiens	RTS beta	2106	96
822	gi12803167	Homo sapiens	nucleosome assembly protein 1-like 1	1728	99
822	gi189067	Homo sapiens	NAP	1728	99
822	gi220496	Mus musculus	nucleosome assembly protein-1	1718	98
823	gi13432042	Homo sapiens	integrin-linked kinase-associated serine/threonine phosphatase 2C	2009	99
823	gi20072498	Mus musculus	Similar to protein phosphatase 2C	1926	94
823	gi3777604	Rattus norvegicus	protein phosphatase 2C	1922	94
824	gi7768636	Xenopus laevis	Kielin	242	36
824	gi6979313	Mus musculus	cysteine-rich repeat-containing protein CRIM1	183	30
824	gi11527817	Homo sapiens	CRIM1 protein	178	30
825	gi21928259	Homo sapiens	seven transmembrane helix receptor	1023	100
825	gi18480746	Mus musculus	olfactory receptor MOR261-10	864	84
825	gi18480744	Mus musculus	olfactory receptor MOR261-9	858	82
826	gi21928655	Homo sapiens	seven transmembrane helix receptor	1458	93
826	gi18480746	Mus musculus	olfactory receptor MOR261-10	1280	79
826	gi18480744	Mus musculus	olfactory receptor MOR261-9	1258	78
827	gi6760369	Mus musculus	ODZ3	364	95

152

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
827	gi4760780	Mus musculus	Ten-m3	364	95
827	gi5307761	Danio rerio	ten-m3	310	78
828	gi21205852	Homo sapiens	T-cell activation Rho GTPase activating protein; TA-GAP	3756	100
828	gi21205854	Homo sapiens	T-cell activation Rho GTPase activating protein splice variant 1; TA-GAP	2850	100
828	gi16265938	Homo sapiens	FKSG15	2439	98
829	gi10432396	Homo sapiens	dJ947L8.1.5 (novel CUB domain protein)	383	62
829	gi14787176	Mus musculus	CSMD1	373	61
829	gi14787181	Homo sapiens	CUB and sushi multiple domains protein 1 short form	369	60
830	gi10432396	Homo sapiens	dJ947L8.1.5 (novel CUB domain protein)	383	62
830	gi14787176	Mus musculus	CSMD1	373	61
830	gi14787181	Homo sapiens	CUB and sushi multiple domains protein 1 short form	369	60
831	gi532124	Dictyostelium discoideum	myosin IC	525	41
831	gi6472600	Chara corallina	unconventional myosin heavy chain	511	43
831	gi9453839	Chara corallina	myosin	511	43
832	gi8953751	Arabidopsis thaliana	myosin heavy chain MYA2	646	40
832	gi6472600	Chara corallina	unconventional myosin heavy chain	646	39
832	gi9453839	Chara corallina	myosin	646	39
833	gi17066528	Canis familiaris	immunoglobulin gamma heavy chain C	42	38
833	gi21113238	Xanthomonas campestris pv. campestris str. ATCC 33913	IS1595 transposase	50	43
833	gi16413516	Listeria innocua	similar to B. subtilis YlaI protein	56	37
834	gi7248845	Homo sapiens	testican-1	2429	99
834	gi793845	Homo sapiens	testican	2429	99
834	gi21265163	Homo sapiens	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican)	2425	99
835	gi12804465	Homo sapiens	prostate cancer overexpressed gene 1	1632	59
835	gi3462515	Homo sapiens	PB39	1632	59
835	gi13111981	Homo sapiens	Similar to selectively expressed in embryonic epithelia protein-1	283	34
836	gi12804465	Homo sapiens	prostate cancer overexpressed gene 1	1637	59
836	gi3462515	Homo sapiens	PB39	1637	59
836	gi13111981	Homo sapiens	Similar to selectively expressed in embryonic epithelia protein-1	283	34
837	gi7689029	Homo sapiens	uncharacterized hypothalamus protein HBEX2	664	100
837	gi17391348	Homo sapiens	Similar to brain expressed, X-linked 1	664	100
837	gi9963771	Homo sapiens	ovarian granulosa cell 13.0 kDa protein hGR74 homolog	664	100
838	gi4585574	Rattus norvegicus	Slit1	287	35
838	gi17380582	Homo sapiens	SLIT1 isoform B	279	35
838	gi4049587	Homo sapiens	Slit-2 protein	297	35
839	gi15488920	Homo sapiens	Similar to RIKEN cDNA 2010107G23 gene	632	100
839	gi19354289	Mus musculus	RIKEN cDNA 2010107G23 gene	570	92
839	gi2267416	Hepatitis D	hepatitis delta antigen	76	33

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		virus			
840	gi21619776	Homo sapiens	Similar to RIKEN cDNA 2600011E07 gene	2491	100
840	gi20988071	Mus musculus	Similar to RIKEN cDNA 2600011E07 gene	921	80
840	gi14531291	Mus musculus	high mobility group protein isoform I	87	34
841	gi21667649	Drosophila melanogaster	myosin binding subunit of myosin phosphatase	231	29
841	gi21392168	Drosophila melanogaster	RE63915p	231	29
841	gi3929221	Homo sapiens	TRF1-interacting ankyrin-related ADP-ribose polymerase	183	32
842	gi12408286	Homo sapiens	apolipoprotein L-IV splice variant a	1742	100
842	gi13374351	Homo sapiens	apolipoprotein L4	1728	99
842	gi12408285	Homo sapiens	apolipoprotein L-IV splice variant b	1683	98
843	gi12408286	Homo sapiens	apolipoprotein L-IV splice variant a	1737	99
843	gi13374351	Homo sapiens	apolipoprotein L4	1723	99
843	gi12408285	Homo sapiens	apolipoprotein L-IV splice variant b	1678	98
844	gi21744725	Homo sapiens	glycosyl-phosphatidyl-inositol-MAM	2296	100
844	gi7529598	Homo sapiens	dJ402N21.3 (novel protein with Immunoglobulin domains)	1048	100
844	gi7529599	Homo sapiens	dJ402N21.1 (novel protein)	662	100
845	gi21744725	Homo sapiens	glycosyl-phosphatidyl-inositol-MAM	5051	100
845	gi7529598	Homo sapiens	dJ402N21.3 (novel protein with Immunoglobulin domains)	1548	99
845	gi7529597	Homo sapiens	dJ402N21.2 (novel protein with MAM domain)	1474	100
846	gi4007758	Schizosaccharomyces pombe	conserved protein; similar to S. cerevisiae YPR144C	633	34
846	gi1066493	Saccharomyces cerevisiae	Weak similarity near C-terminus to RNA Polymerase beta subunit (Swiss Prot. accession number P11213) and CCAAT-binding transcription factor (PIR accession number A36368)	482	32
846	gi18086412	Arabidopsis thaliana	At2g17250/T23A1.11	420	44
847	gi14701768	Homo sapiens	Vam6/Vps39-like protein	3499	96
847	gi14280050	Homo sapiens	Vps39/Vam6-like protein	3499	96
847	gi18857927	Mus musculus	VPS39 long isoform	3409	93
848	gi3811347	Homo sapiens	cytosolic phospholipase A2 beta	1209	44
848	gi4886978	Homo sapiens	cytosolic phospholipase A2 beta; cPLA2beta	1209	44
848	gi190004	Homo sapiens	phosphatidylcholine 2-acylhydrolase	512	35
849	gi7291437	Drosophila melanogaster	CG4071-PA	516	51
849	gi17946619	Drosophila melanogaster	RH31535p	217	42
849	gi21645615	Drosophila melanogaster	CG4071-PB	217	42
850	gi13161409	Mus musculus	family 4 cytochrome P450	444	73
850	gi5263306	Coptotermes acinaciformis	family 4 cytochrome P450	200	41
850	gi13182964	Mus musculus	cytochrome P450 CYP4F13	196	38
851	gi13447749	Homo sapiens	fibroblast growth factor receptor 5	2475	98
851	gi10944887	Homo sapiens	FGFR-like protein	2475	98

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
851	gi13183618	Homo sapiens	FGF homologous factor receptor	2421	97
852	gi13447749	Homo sapiens	fibroblast growth factor receptor 5	2701	99
852	gi10944887	Homo sapiens	FGFR-like protein	2701	99
852	gi13183618	Homo sapiens	FGF homologous factor receptor	2647	98
853	gi13183618	Homo sapiens	FGF homologous factor receptor	583	98
853	gi13447749	Homo sapiens	fibroblast growth factor receptor 5	583	98
853	gi10944887	Homo sapiens	FGFR-like protein	583	98
854	gi643656	Rattus norvegicus	synaptotagmin VII	2035	95
854	gi12667446	Rattus norvegicus	synaptotagmin VIIs	2035	95
854	gi6136786	Mus musculus	synaptotagmin VII	2026	95
855	gi12053709	Homo sapiens	a disintegrin-like and metalloprotease (repolysin type) with thrombospondin type 1 motif, 12	8842	100
855	gi5923788	Homo sapiens	zinc metalloprotease ADAMTS7	2489	58
855	gi19171178	Homo sapiens	metalloprotease disintegrin 16 with thrombospondin type I motif	1598	39
856	gi15929988	Homo sapiens	Similar to TLH29 protein precursor	155	86
856	gi7649139	Homo sapiens	pIFI27-like protein	83	44
856	gi11493982	Homo sapiens	TLH29 protein precursor	83	44
857	gi13542874	Mus musculus	Similar to CGI-67 protein	1299	74
857	gi21707079	Homo sapiens	similar to RIKEN cDNA 2210412D01	1278	75
857	gi4929603	Homo sapiens	CGI-67 protein	1087	81
858	gi13542874	Mus musculus	Similar to CGI-67 protein	1299	74
858	gi21707079	Homo sapiens	similar to RIKEN cDNA 2210412D01	1279	73
858	gi4929603	Homo sapiens	CGI-67 protein	1087	81
859	gi21595166	Mus musculus	RIKEN cDNA 4933425F03 gene	1823	83
859	gi16359267	Mus musculus	Similar to RIKEN cDNA 4933425F03 gene	1822	83
859	gi21619888	Homo sapiens	Similar to RIKEN cDNA 4933425F03 gene	1542	98
860	gi21595166	Mus musculus	RIKEN cDNA 4933425F03 gene	2278	88
860	gi16359267	Mus musculus	Similar to RIKEN cDNA 4933425F03 gene	2277	88
860	gi21619888	Homo sapiens	Similar to RIKEN cDNA 4933425F03 gene	1958	99
861	gi11493463	Homo sapiens	PRO2852	301	75
861	gi14189960	Homo sapiens	PRO0764	271	65
861	gi21104464	Homo sapiens	OK/SW-CL.41	264	70
863	gi21320872	Mus musculus	Cog8	2747	88
863	gi17862986	Drosophila melanogaster	SD07339p	795	45
863	gi5922593	Schizosaccharomyces pombe	pi008	230	21
864	gi21618851	Mus musculus	RIKEN cDNA 2610510L01 gene	882	92
864	gi20977573	Danio rerio	U1 small nuclear ribonucleoprotein C	75	32
864	gi1562574	Mus musculus	U1 snRNP-specific protein C	75	32
865	gi17862312	Drosophila melanogaster	LD21841p	646	41
865	gi22294210	Thermosynechococcus elongatus BP-1	WD-40 repeat protein	123	27
865	gi886024	Thermomonospora	PkwA	124	25

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		ora curvata			
866	gi3878846	Caenorhabditis elegans	R05D7.3	119	37
866	gi1685056	Xenopus laevis	Pax6	87	24
866	gi8132389	Xenopus laevis	paired domain transcription factor variant A	81	23
867	gi12406973	Homo sapiens	alanine-glyoxylate aminotransferase 2	2740	100
867	gi1944136	Rattus norvegicus	beta-alanine-pyruvate aminotransferase	2255	83
867	gi1000448	Rattus norvegicus	Rat kidney AGT2 precursor	2208	81
868	gi12406973	Homo sapiens	alanine-glyoxylate aminotransferase 2	1870	98
868	gi1944136	Rattus norvegicus	beta-alanine-pyruvate aminotransferase	1630	86
868	gi1000448	Rattus norvegicus	Rat kidney AGT2 precursor	1583	84
869	gi4165315	Sus scrofa	kallikrein	468	42
869	gi190263	Homo sapiens	plasma prekallikrein	467	38
869	gi8809781	Homo sapiens	plasma kallikrein precursor	467	38
870	gi17985046	Brucella melitensis	GLYCOSYL TRANSFERASE	137	28
870	gi5478237	Brucella melitensis	Bme7	137	28
870	gi20906785	Methanosarcina mazei Goel	Transposase	126	25
871	gi4565840	Cnemidophorus tigris	cytochrome b oxidase	76	41
871	gi15023030	Clostridium acetobutylicum	Uncharacterized membrane protein, ortholog YYAS B.subtilis	72	44
871	gi7549241	Barbatia tenera	cytochrome oxidase subunit I	71	28
872	gi8705222	Homo sapiens	IL-17B receptor	1998	100
872	gi9246433	Homo sapiens	IL-17 receptor homolog precursor	1996	99
872	gi9246429	Mus musculus	IL-17 receptor homolog precursor	1504	75
873	gi18676472	Homo sapiens	FLJ00133 protein	6475	100
873	gi18676498	Homo sapiens	FLJ00146 protein	2352	100
873	gi161467	Strongylocentrotus purpuratus	fibropellin Ia	1246	38
874	gi213198	Petromyzon marinus	fibrinogen alpha chain	89	39
874	gi15292317	Drosophila melanogaster	LD46863p	87	34
874	gi4877921	Streptococcus pyogenes	serum opacity factor precursor	81	33
875	gi14249936	Homo sapiens	Similar to S-adenosylhomocysteine hydrolase-like 1	2582	97
875	gi17390493	Mus musculus	S-adenosylhomocysteine hydrolase-like 1	2429	92
875	gi2852125	Homo sapiens	S-adenosyl homocysteine hydrolase homolog	2429	92
876	gi14279990	Homo sapiens	ubiquitin UBF-fl	458	100
876	gi6706799	Homo sapiens	dJ447F3.2.1 (ubiquitin-conjugating enzyme E2 H10 (isoform 1))	214	74
876	gi14043322	Homo sapiens	ubiquitin carrier protein E2-C	214	74
877	gi20086516	Homo sapiens	prominin-related protein	4241	99
877	gi20086520	Mus musculus	prominin-related protein	3157	73

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
877	gi19909067	Rattus norvegicus	testosterone-regulated prominin-related protein	2920	69
878	gi13159480	Homo sapiens	Translation may initiate at the ATG codon at nucleotides 40-42 or the ATG at nucleotides 43-45	2104	100
878	gi21483846	Sus scrofa	fibrinogen-like protein 2	406	36
878	gi9229906	Ciona intestinalis	fibrinogen-like protein	408	36
879	gi13159480	Homo sapiens	Translation may initiate at the ATG codon at nucleotides 40-42 or the ATG at nucleotides 43-45	2100	99
879	gi21483846	Sus scrofa	fibrinogen-like protein 2	406	36
879	gi9229906	Ciona intestinalis	fibrinogen-like protein	408	36
880	gi13159480	Homo sapiens	Translation may initiate at the ATG codon at nucleotides 40-42 or the ATG at nucleotides 43-45	2100	99
880	gi21483846	Sus scrofa	fibrinogen-like protein 2	406	36
880	gi9229906	Ciona intestinalis	fibrinogen-like protein	408	36
881	gi11493483	Homo sapiens	PRO2550	322	66
881	gi7770139	Homo sapiens	PRO1722	318	69
881	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	304	72
882	gi10175777	Bacillus halodurans	protease specific for phage lambda cII repressor	67	34
882	gi15558903	Xenopus laevis	Tob	64	51
882	gi21998835	Rattus norvegicus	monocarboxylate transporter 8	67	33
883	gi18073362	Homo sapiens	cystine/glutamate transporter	2552	100
883	gi11493652	Homo sapiens	calcium channel blocker resistance protein CCB1	2552	100
883	gi13924720	Homo sapiens	cystine/glutamate transporter xCT	2552	100
884	gi507213	Homo sapiens	serine kinase	1797	97
884	gi14252988	Homo sapiens	SRPK1a protein kinase	1797	97
884	gi3135975	Homo sapiens	dJ422H11.1.1 (Serine Kinase) (isoform 1)	1796	98
885	gi9837288	Homo sapiens	C-type lectin	271	54
885	gi6651065	Homo sapiens	lectin-like NK cell receptor LLT1	271	54
885	gi18044358	Homo sapiens	Similar to lectin-like NK cell receptor	270	57
886	gi22164066	Homo sapiens	neuroblastoma-amplified protein	7571	99
886	gi5833317	Oryzias latipes	mixed lineage leukemia-like protein	89	23
886	gi7108717	Nicotiana tabacum	MAR-binding protein MFP1 homolog	89	31
887	gi22164066	Homo sapiens	neuroblastoma-amplified protein	6897	98
887	gi5833317	Oryzias latipes	mixed lineage leukemia-like protein	89	23
888	gi17430957	Ralstonia solanacearum	HYPOTHETICAL TRANSMEMBRANE PROTEIN	453	40
888	gi13421965	Caulobacter crescentus CB15	M20/M25/M40 family peptidase	377	38
888	gi2330791	Schizosaccharomyces pombe	carboxypeptidase s precursor	352	33
889	gi11558029	Homo sapiens	organic cation transporter	1860	99
889	gi18088251	Homo sapiens	Similar to hBOIT for potent brain type organic ion transporter	1206	97

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
889	gi9663117	Homo sapiens	organic cation transporter	1852	99
890	gi344112	synthetic construct	chloramphenicol acetyltransferase and carboxy terminal fusion protein	57	28
890	gi412284	synthetic construct	carboxy terminal fusion protein	57	28
890	gi13122523	Barbus brachycephalus	ATP synthase 8	56	28
891	gi13375149	Homo sapiens	dJ1118M15.2 (Novel protein)	538	98
891	gi7259265	Mus musculus	contains transmembrane (TM) region	269	48
891	gi1806278	Rattus norvegicus	glycoprotein 56	143	35
892	gi16589003	Homo sapiens	bromodomain-containing 4	6353	99
892	gi9931486	Mus musculus	cell proliferation related protein CAP	5635	90
892	gi18308125	Mus musculus	bromodomain-containing protein BRD4 long variant	5633	90
893	gi15420828	Homo sapiens	NOE3-1	2504	99
893	gi19386926	Rattus norvegicus	optimedlin form B	2484	98
893	gi19386930	Mus musculus	optimedlin form B	2484	98
894	gi10336599	Xenopus laevis	folliculin-related protein	234	32
894	gi349006	Mus musculus	TGF-beta-inducible protein	225	29
894	gi20810033	Mus musculus	folliculin-like	223	29
895	gi5002565	Takifugu rubripes	cysteine conjugate beta-lyase	1244	55
895	gi758591	Homo sapiens	glutamine--phenylpyruvate aminotransferase	1201	51
895	gi15425868	Aedes aegypti	kynurenine aminotransferase	1188	55
896	gi20522012	Homo sapiens	similar to an actin bundling protein, dematn.	1312	57
896	gi2337952	Homo sapiens	actin-binding double-zinc-finger protein	1312	57
896	gi21666433	Mus musculus	actin-binding LIM protein 1 medium isoform	1305	57
898	gi6716518	Mus musculus	doublecortin-like kinase	821	52
898	gi21619202	Homo sapiens	Similar to doublecortin and CaM kinase-like 1	810	51
898	gi20152113	Drosophila melanogaster	RE56868p	778	45
899	gi9280108	Macaca fascicularis	membrane-associated prostaglandin E synthase-2	1907	97
899	gi9757960	Arabidopsis thaliana	contains similarity to glutathione-S-transferase/glutaredoxin~gene_id:MJC20.26	396	50
899	gi17944528	Drosophila melanogaster	RH17614p	566	42
900	gi4894854	Homo sapiens	complement C1q A chain precursor	1308	99
900	gi20988805	Homo sapiens	complement component 1, q subcomponent, alpha polypeptide	1308	99
900	gi12805247	Mus musculus	complement component 1, q subcomponent, alpha polypeptide	945	70
901	gi10176989	Arabidopsis thaliana	contains similarity to hedgehog-interacting protein~gene_id:MYH19.17	86	34
901	gi456384	Blastocrithidia culicis	apocytochrome B	41	50
902	gi2565046	Homo sapiens	CAGF28	3775	97
902	gi21707458	Homo sapiens	PAX transcription activation domain	2709	87

158

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
			interacting protein 1 like		
902	gi4336734	Mus musculus	Pax transcription activation domain interacting protein PTIP	2473	80
903	gi4336734	Mus musculus	Pax transcription activation domain interacting protein PTIP	531	93
903	gi14164561	Xenopus laevis	Swift	467	79
903	gi12382298	Human herpesvirus 8	OrfK10	48	34
904	gi19353375	Mus musculus	RIKEN cDNA 1110031I02 gene	745	78
904	gi15929776	Homo sapiens	growth suppressor 1	137	41
904	gi5805194	Rattus norvegicus	leprecan	137	41
905	gi2443352	Mus musculus	platelet glycoprotein Ib beta	150	45
905	gi21355064	Homo sapiens	platelet glycoprotein Ib beta chain	146	43
905	gi306792	Homo sapiens	platelet glycoprotein Ib beta chain precursor	146	43
906	gi13991166	Homo sapiens	sialic acid-binding immunoglobulin-like lectin-like short splice variant	1174	100
906	gi13991167	Homo sapiens	sialic acid-binding immunoglobulin-like lectin-like long splice variant	1174	100
906	gi14625822	Homo sapiens	Siglec-L1	1174	100
907	gi21708018	Mus musculus	RIKEN cDNA 2700029E10 gene	626	66
907	gi7547035	Homo sapiens	SGC32445 protein	474	63
907	gi21626575	Drosophila melanogaster	CG30193-PA	457	55
908	gi6273399	Homo sapiens	melanoma-associated antigen MG50	2748	60
908	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	2748	60
908	gi531385	Drosophila melanogaster	peroxidase precursor	1721	42
909	gi6273399	Homo sapiens	melanoma-associated antigen MG50	2748	60
909	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	2748	60
909	gi531385	Drosophila melanogaster	peroxidase precursor	1721	42
910	gi6273399	Homo sapiens	melanoma-associated antigen MG50	2799	59
910	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	2799	59
910	gi531385	Drosophila melanogaster	peroxidase precursor	1708	41
911	gi18182323	Mus musculus	crumbs-like protein 1 precursor	777	31
911	gi6014482	Homo sapiens	CRB1	754	30
911	gi18175289	Homo sapiens	CRB1 isoform I precursor	754	30
912	gi6650802	Homo sapiens	PRO1848	205	56
912	gi21104464	Homo sapiens	OK/SW-CL.41	188	61
912	gi11493463	Homo sapiens	PRO2852	175	54
913	gi6808611	Homo sapiens	88-kDa Golgi protein	3237	99
913	gi6969980	Homo sapiens	golgin 67	2345	98
913	gi7211438	Homo sapiens	golgin-67	2330	98
914	gi307377	Homo sapiens	cAMP-dependent protein kinase RI-beta regulatory subunit	1957	99
914	gi200365	Mus musculus	cAMP-dependent protein kinase regulatory subunit	1886	94
914	gi15030299	Mus musculus	Similar to protein kinase, cAMP	1881	94

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
			dependent regulatory, type I beta		
915	gi20306468	Mus musculus	Similar to RIKEN cDNA 2610025P08 gene	382	41
915	gi7161798	Homo sapiens	dJ470B24.1.1 (myeloid/lymphoid or mixed-lineage leukemia (trithorax (Drosophila) homolog); translocated to, 4 (AF-6) (isoform 1))	130	32
915	gi7161797	Homo sapiens	dJ470B24.1.2 (myeloid/lymphoid or mixed-lineage leukemia (trithorax (Drosophila) homolog); translocated to, 4 (AF-6) (isoform 2))	130	32
916	gi1845577	Mus musculus	arachidonate 12(S)-lipoxygenase	2633	77
916	gi3645913	Mus musculus	12(S)-lipoxygenase	2633	77
916	gi15489302	Mus musculus	Similar to arachidonate 15-lipoxygenase	2631	77
917	gi15489302	Mus musculus	Similar to arachidonate 15-lipoxygenase	751	78
917	gi1845577	Mus musculus	arachidonate 12(S)-lipoxygenase	748	78
917	gi1101886	Mus musculus	arachidonate lipoxygenase	748	78
918	gi15489302	Mus musculus	Similar to arachidonate 15-lipoxygenase	1266	75
918	gi1845577	Mus musculus	arachidonate 12(S)-lipoxygenase	1263	75
918	gi1101886	Mus musculus	arachidonate lipoxygenase	1263	75
919	gi13661964	Leishmania major	L344.3	108	21
919	gi17135639	Nostoc sp. PCC 7120	WD-repeat protein	95	21
919	gi11139242	Homo sapiens	meiotic recombination protein REC14	93	25
920	gi17862298	Drosophila melanogaster	LD21662p	627	42
920	gi2425111	Dictyostelium discoideum	ZipA	122	28
920	gi641958	Homo sapiens	non-muscle myosin B	118	24
921	gi8132683	Homo sapiens	cytokine-like protein C17	241	64
921	gi12751073	Homo sapiens	PNAS-31	74	92
921	gi11323101	Saint Croix river virus	VP4	79	32
922	gi8132683	Homo sapiens	cytokine-like protein C17	241	64
922	gi12751073	Homo sapiens	PNAS-31	74	92
922	gi11323101	Saint Croix river virus	VP4	79	32
923	gi8132683	Homo sapiens	cytokine-like protein C17	384	73
923	gi12751073	Homo sapiens	PNAS-31	74	92
923	gi216168	Bacteriophage SPP1	promoter 3 protein	56	37
924	gi8132683	Homo sapiens	cytokine-like protein C17	263	98
924	gi1143067	Canis familiaris	alpha-L-fucosidase	69	59
924	gi309444	Mus musculus	MRK	58	65
925	gi8132683	Homo sapiens	cytokine-like protein C17	591	100
925	gi3406819	Mus musculus	growth factor receptor	64	60
925	gi12724591	Lactococcus lactis subsp. lactis	UNKNOWN PROTEIN	41	37
926	gi17975777	Homo sapiens	vesicular inhibitory amino acid transporter	2741	99
926	gi13396317	Homo sapiens	bA122O1.1 (A novel protein (ortholog of the mouse vesicular inhibitory amino acid transporter, VIAAT))	2741	99

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
926	gi2587061	Rattus norvegicus	vesicular GABA transporter	2694	98
927	gi3097285	Rattus norvegicus	ZOG	670	39
927	gi802014	Rattus norvegicus	preadipocyte factor 1	665	39
927	gi13365691	Mus musculus	dlk (Delta like)	649	39
928	gi6624073	Homo sapiens	similar to hepatitis delta antigen interacting protein A ; similar to AAB05928.1 (PID:g1488314)	1757	93
928	gi1488314	Homo sapiens	hepatitis delta antigen interacting protein A	274	45
928	gi16768374	Drosophila melanogaster	GM03282p	359	37
929	gi4337106	Homo sapiens	BAT4	864	98
929	gi14250638	Homo sapiens	Similar to DNA segment, Chr 17, human D6S54E	864	98
929	gi3941733	Mus musculus	BAT4	581	71
930	gi9759107	Arabidopsis thaliana	phosphate/phosphoenolpyruvate translocator protein-like	289	30
930	gi21536504	Arabidopsis thaliana	phosphate/phosphoenolpyruvate translocator-like protein	245	27
930	gi8778643	Arabidopsis thaliana	F5O11.25	235	29
931	gi5852981	Homo sapiens	cardiotrophin-like cytokine CLC	1204	99
931	gi6007641	Homo sapiens	neurotrophin-1/B-cell stimulating factor-3	1204	99
931	gi15277895	Homo sapiens	Similar to cardiotrophin-like cytokine; neurotrophin-1/B-cell stimulating factor-3	1204	99
932	gi22003732	Homo sapiens	MTLC	853	99
932	gi18490933	Homo sapiens	Similar to RIKEN cDNA 1110020B04 gene	846	98
932	gi20453974	Mus musculus	MT-MC1	718	82
933	gi9958075	Arabidopsis thaliana	Putative methionine aminopeptidase	739	53
933	gi11320956	Arabidopsis thaliana	methionine aminopeptidase-like protein	739	53
933	gi21553973	Arabidopsis thaliana	methionyl aminopeptidase-like protein	717	52
934	gi4104963	Rattus norvegicus	neurexophilin 4	1493	90
934	gi1336013	Mus musculus	neurexophilin 2	327	65
934	gi4105164	Homo sapiens	neurexophilin 2	323	65
935	gi15025812	Clostridium acetobutylicum	Methyl-accepting chemotaxis protein with HAMP domain	65	38
935	gi17224936	Trypanosoma brucei	corset-associated protein 15	63	31
935	gi15025892	Clostridium acetobutylicum	Ribosome-associated protein Y (PSrp-1)	48	38
936	gi16197625	Arabidopsis thaliana	anaphase promoting complex subunit 11	64	32
936	gi10834682	Homo sapiens	PP3958	74	46
937	gi19387136	Homo sapiens	PYRIN-containing APAF1-like protein 5	874	99
937	gi202806	Rattus norvegicus	vasopressin receptor	561	68
937	gi21410402	Mus musculus	expressed sequence AI504961	532	67

161

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
938	gi11321325	Homo sapiens	Lin-7b	1030	100
938	gi20381193	Homo sapiens	Lin-7b protein; likely ortholog of mouse LIN-7B; mammalian LIN-7 protein 2	1030	100
938	gi3885828	Rattus norvegicus	lin-7-A	1019	98
939	gi14349125	Homo sapiens	alpha2-glucosyltransferase	738	96
939	gi3513451	Rattus norvegicus	potassium channel regulator 1	718	93
939	gi21711799	Drosophila melanogaster	RH44301p	142	32
940	gi12803183	Homo sapiens	polypyrimidine tract binding protein (heterogeneous nuclear ribonucleoprotein I)	1527	91
940	gi32354	Homo sapiens	nuclear ribonucleoprotein	1527	91
940	gi35772	Homo sapiens	polypirimidine tract binding protein	1527	91
941	gi6752658	Homo sapiens	epidermal growth factor repeat containing protein	3046	99
941	gi16040981	Mus musculus	POEM	884	51
941	gi15430246	Mus musculus	nephronectin short isoform	884	51
942	gi6752658	Homo sapiens	epidermal growth factor repeat containing protein	3036	98
942	gi16040981	Mus musculus	POEM	884	51
942	gi15430246	Mus musculus	nephronectin short isoform	884	51
943	gi17980969	Homo sapiens	se14-3r protein	5146	99
943	gi11385648	Homo sapiens	CTCL tumor antigen se14-3	3867	99
943	gi7960216	Homo sapiens	RACK-like protein PRKCBP1	3124	99
944	gi17980969	Homo sapiens	se14-3r protein	3140	99
944	gi13677201	Homo sapiens	dJ569M23.1.2 (protein kinase C binding protein 1, isoform 2)	2771	100
944	gi13677198	Homo sapiens	dJ569M23.1.3 (protein kinase C binding protein 1, isoform 3 (DKFZp564P1772))	2638	96
945	gi17980969	Homo sapiens	se14-3r protein	3550	84
945	gi13677201	Homo sapiens	dJ569M23.1.2 (protein kinase C binding protein 1, isoform 2)	2771	100
945	gi13677198	Homo sapiens	dJ569M23.1.3 (protein kinase C binding protein 1, isoform 3 (DKFZp564P1772))	2638	96
946	gi17980969	Homo sapiens	se14-3r protein	3550	84
946	gi13677198	Homo sapiens	dJ569M23.1.3 (protein kinase C binding protein 1, isoform 3 (DKFZp564P1772))	2380	90
946	gi13677201	Homo sapiens	dJ569M23.1.2 (protein kinase C binding protein 1, isoform 2)	2377	90
947	gi14043211	Homo sapiens	Similar to RIKEN cDNA 4931428F04 gene	2410	98
947	gi22204070	Macaca mulatta	metabotropic glutamate receptor 1	91	42
947	gi170454	Lycopersicon esculentum	cell wall hydroxyproline-rich glycoprotein	70	39
948	gi14972753	Streptococcus pneumoniae TIGR4	alcohol dehydrogenase, zinc-containing	51	33
948	gi20152351	Avian infectious bronchitis virus	spike glycoprotein S1 subunit	68	34
948	gi9658106	Vibrio cholerae	polyhydroxyalkanoic acid synthase	67	26
949	gi19387136	Homo sapiens	PYRIN-containing APAF1-like protein 5	1738	99
949	gi202806	Rattus	vasopressin receptor	1037	64

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		norvegicus			
949	gi21410402	Mus musculus	expressed sequence AI504961	988	63
950	gi3978472	Rattus norvegicus	potassium channel subunit	5393	88
950	gi20338417	Gallus gallus	potassium channel subunit	4792	88
950	gi7303760	Drosophila melanogaster	CG12904-PA	981	62
951	gi18147612	Homo sapiens	metalloprotease disintegrin	3535	99
951	gi21908028	Homo sapiens	a disintegrin and metalloprotease domain 33	3535	99
951	gi13157560	Homo sapiens	dJ964F7.1 (novel disintegrin and reprotolysin metalloproteinase family protein)	3078	99
952	gi18606367	Mus musculus	RIKEN cDNA 4930570C03 gene	715	92
952	gi9971130	Schizosaccharomyces pombe	human downs syndrome critical region-like	72	31
952	gi5708224	Rhodoblastus acidophilus	LH2alpha5	60	31
953	gi15420879	Mus musculus	ankyrin repeat-containing SOCS box protein 10	2053	82
953	gi18092200	Homo sapiens	ASB-10	1909	98
953	gi18031949	Mus musculus	SOCS box protein ASB-18	816	45
954	gi491284	synthetic construct	IFN-pseudo-omega 2	799	98
954	gi386800	Homo sapiens	interferon-alpha	330	72
954	gi490110	Homo sapiens	interferon-omega 1	330	72
955	gi9844580	Homo sapiens	dJ1153D9.4 (novel protein)	623	84
955	gi9844579	Homo sapiens	dJ1153D9.3 (novel protein)	450	97
955	gi15928971	Homo sapiens	Similar to neuronal thread protein	430	90
956	gi12804321	Homo sapiens	peroxisomal short-chain alcohol dehydrogenase	685	100
956	gi19113668	Homo sapiens	NADP-dependent retinol dehydrogenase short isoform	878	100
956	gi11559412	Homo sapiens	NADPH-dependent retinol dehydrogenase/reductase	587	100
957	gi12718818	Mus musculus	sulfhydryl oxidase	496	49
957	gi12718820	Rattus norvegicus	sulfhydryl oxidase	489	47
957	gi12483919	Rattus norvegicus	FAD-dependent sulfhydryl oxidase-2	489	47
958	gi12958660	Homo sapiens	acid phosphatase	2252	100
958	gi12958663	Homo sapiens	acid phosphatase variant 3	1285	99
958	gi52871	Mus musculus	lysosomal acid phosphatase	837	45
959	gi28966	Homo sapiens	alpha 1-antitrypsin	1703	100
959	gi6855601	Homo sapiens	PRO0684	1703	100
959	gi11493443	Homo sapiens	PRO2209	1703	100
960	gi28966	Homo sapiens	alpha 1-antitrypsin	1080	100
960	gi11493443	Homo sapiens	PRO2209	1080	100
960	gi177829	Homo sapiens	alpha-1-antitrypsin	1080	100
961	gi28966	Homo sapiens	alpha 1-antitrypsin	1239	100
961	gi11493443	Homo sapiens	PRO2209	1239	100
961	gi177829	Homo sapiens	alpha-1-antitrypsin	1239	100
962	gi28966	Homo sapiens	alpha 1-antitrypsin	1574	93
962	gi11493443	Homo sapiens	PRO2209	1574	93

163

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
962	gi177829	Homo sapiens	alpha-1-antitrypsin	1574	93
963	gi6706993	Streptomyces coelicolor A3(2)	methyltransferase	83	26
963	gi7303904	Drosophila melanogaster	CG13954-PA	85	53
964	gi2632092	Pongo pygmaeus	fertilin alpha protein	4128	92
964	gi794073	Macaca fascicularis	fertilin alpha-I	3136	74
964	gi1841702	Macaca fascicularis	fertilin alpha-I isoform	3136	74
965	gi4107229	Homo sapiens	lipophilin A	454	100
965	gi4107231	Homo sapiens	lipophilin B	267	60
965	gi17887359	Oryctolagus cuniculus	lipophilin AL2	248	54
966	gi3335100	Homo sapiens	CD39L3	2816	100
966	gi13817037	Homo sapiens	E-type ATPase	2812	99
966	gi20988653	Homo sapiens	Similar to ectonucleoside triphosphate diphosphohydrolase 3	2413	99
967	gi6942096	Mus musculus	CBLN3	936	93
967	gi180251	Homo sapiens	precerebellin	549	57
967	gi5702371	Mus musculus	precerebellin-1	542	56
968	gi17390957	Mus musculus	Similar to RIKEN cDNA 2010001E11 gene	129	32
968	gi16410838	Listeria monocytogenes	similar to multidrug-efflux transporter	95	27
968	gi4914624	Listeria monocytogenes	multidrug resistance transporter	95	27
969	gi17390957	Mus musculus	Similar to RIKEN cDNA 2010001E11 gene	191	26
969	gi2828808	Bacillus subtilis	glucose transporter	100	23
969	gi14023148	Mesorhizobium loti	probable fosmidomycin resistance protein	112	25
970	gi13161123	Homo sapiens	transcript Y 10	151	54
970	gi4545317	Acipenser ruthenus	immunoglobulin light chain precursor	160	25
970	gi9937599	Salmo trutta	MHC class I heavy chain	160	31
971	gi4160197	Homo sapiens	dJ327J16.2 (supported by GENSCAN and GENEWISE)	2515	99
971	gi2253263	Rattus norvegicus	neuronal pentraxin receptor	2238	89
971	gi12744624	Mus musculus	neuronal pentraxin receptor	2212	88
972	gi4760782	Mus musculus	Ten-m4	4188	96
972	gi3170615	Mus musculus	DOC4	4166	96
972	gi5307785	Danio rerio	ten-m4	3537	78
973	gi14714932	Homo sapiens	Similar to nuclear factor (erythroid-derived 2)-like 1	3770	100
973	gi473090	Mus musculus	NFE2-related factor 1	3644	96
973	gi3978250	Mus musculus	Nrfl splice variant D	3280	96
974	gi7716100	Rattus norvegicus	selective LIM binding factor	8413	95
974	gi17044301	Leishmania major	possible LIM-binding factor	2139	36
974	gi10440379	Homo sapiens	FLJ00025 protein	135	25

164

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
975	gi20799661	Mus musculus	mucolipin-2	1593	72
975	gi20987535	Mus musculus	RIKEN cDNA 3300002C04 gene	1590	71
975	gi19072756	Mus musculus	mucolipin-3	1136	51
976	gi20799661	Mus musculus	mucolipin-2	2394	83
976	gi20987535	Mus musculus	RIKEN cDNA 3300002C04 gene	2391	82
976	gi19072754	Homo sapiens	mucolipin-3	1674	59
977	gi403020	Mus musculus	En-2/lacZ fusion protein	988	96
977	gi14193747	Mus musculus	zinc finger 142	258	24
977	gi1510147	Homo sapiens	similar to Human zinc finger protein(ZNF142)	223	20
978	gi10581238	Halobacterium sp. NRC-1	Vng1783h	54	46
978	gi19699294	Arabidopsis thaliana	AT3g48750/T21J18_20	73	30
979	gi7959724	Homo sapiens	PRO0929	63	30
979	gi13540242	Anopheles stephensi	NADH dehydrogenase subunit 5	62	31
979	gi20904847	Methanosarcina mazei Goel	8-oxoguanine DNA glycosylase	64	40
980	gi5281519	Homo sapiens	HTRA serine protease	2164	100
980	gi1513059	Homo sapiens	serin protease with IGF-binding motif	2164	100
980	gi1621244	Homo sapiens	novel serine protease, PRSS11	2164	100
981	gi7008025	Callithrix jacchus	prochymosin	832	68
981	gi19851892	Bos taurus	chymosin precursor	515	77
981	gi162860	Bos taurus	preprochymosin b	752	62
982	gi18461371	Rattus norvegicus	sulfatase FP	276	68
982	gi21961489	Mus musculus	Similar to sulfatase FP	276	68
982	gi15430244	Coturnix coturnix	N-acetylglucosamine-6-sulfatase	263	68
983	gi3043872	Lactococcus lactis	transmembrane protein Tmp3	69	32
983	gi17428881	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	62	34
983	gi433707	Zea mays	prolin rich protein	63	48
984	gi6013463	Bothrops jararaca	carboxypeptidase homolog	826	46
984	gi9558448	Mus musculus	carboxypeptidase R	812	45
984	gi7416967	Mus musculus	thrombin-activatable fibrinolysis inhibitor	812	45
985	gi6013463	Bothrops jararaca	carboxypeptidase homolog	826	46
985	gi9558448	Mus musculus	carboxypeptidase R	812	45
985	gi7416967	Mus musculus	thrombin-activatable fibrinolysis inhibitor	812	45
986	gi11545707	Homo sapiens	ISCU2	845	100
986	gi20381021	Mus musculus	RIKEN cDNA 2310020H20 gene	807	96
986	gi11545705	Homo sapiens	ISCU1	663	99
987	gi12314022	Homo sapiens	dJ553F4.4 (Novel protein similar to Drosophila CG8055 protein)	881	89
987	gi22417143	Homo sapiens	CGI-301 protein	853	100
987	gi13182765	Homo sapiens	CDA04	560	60
988	gi52959	Mus musculus	precursor polypeptide (AA -26 to 108)	146	34
988	gi198922	Mus musculus	lymphocyte differentiation antigen	145	34
988	gi198926	Mus musculus	Ly-6A.2 alloantigen	145	34

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
990	gi15990480	Homo sapiens	Similar to AE-binding protein 2	1570	100
990	gi4106464	Mus musculus	AE-1 binding protein AEBP2	1555	98
990	gi21595036	Mus musculus	AE binding protein 2	1555	98
991	gi23903	Homo sapiens	63kDa protein kinase	2897	99
991	gi204058	Rattus norvegicus	extracellular signal-related kinase 3	1499	62
991	gi16306437	Homo sapiens	ERK-3	1492	62
992	gi17016967	Homo sapiens	NUANCE	3403	90
992	gi17861384	Homo sapiens	nesprin-2 gamma	3403	90
992	gi21748548	Homo sapiens	FLJ00347 protein	3403	90
993	gi20070711	Homo sapiens	similar to RIKEN cDNA 2310044D20	997	100
993	gi18204756	Mus musculus	Similar to RIKEN cDNA 2310044D20 gene	626	68
993	gi7304139	Drosophila melanogaster	CG12159-PA	111	28
994	gi14278927	Mus musculus	gliacolin	866	68
994	gi10566471	Mus musculus	Gliacolin	866	68
994	gi3747099	Mus musculus	C1q-related factor	734	67
995	gi20987689	Homo sapiens	Similar to allantoicase	1838	99
995	gi14718648	Homo sapiens	allantoicase	1633	99
995	gi9255889	Mus musculus	allantoicase	1476	77
997	gi2522208	Homo sapiens	Ras-GRF2	6407	99
997	gi5882290	Homo sapiens	Ras guanine nucleotide exchange factor 2	6401	99
997	gi57665	Rattus rattus	P140 RAS-GRF	4121	65
998	gi22038159	Homo sapiens	zizimin1	8544	100
998	gi14597976	Homo sapiens	human CLASP-4	3533	56
998	gi550420	Rattus norvegicus	trg	2842	87
999	gi17861850	Drosophila melanogaster	GM03763p	334	70
999	gi17862036	Drosophila melanogaster	LD05823p	265	47
999	gi10178624	Mus musculus	SETA binding protein 1; SB1	215	45
1000	gi21594273	Homo sapiens	SAC2 suppressor of actin mutations 2-like (yeast)	3626	100
1000	gi14041697	Homo sapiens	dJ1033B10.5.1 (SAC2 (suppressor of actin mutations 2, yeast, homolog)-like (ARE1), isoform 1)	3587	99
1000	gi3850063	Rattus norvegicus	ARE1	3576	98
1001	gi1438534	Rattus norvegicus	rA9	4002	61
1001	gi1438532	Rattus norvegicus	rA1	430	36
1001	gi9438033	Homo sapiens	ser/arg-rich pre-mRNA splicing factor SR-A1	407	35
1002	gi1438534	Rattus norvegicus	rA9	4002	61
1002	gi9438033	Homo sapiens	ser/arg-rich pre-mRNA splicing factor SR-A1	407	35
1002	gi10440402	Homo sapiens	FLJ00034 protein	407	35
1003	gi1675220	Cricetulus griseus	SREBP cleavage activating protein	6200	92
1003	gi20378357	Drosophila	ER-golgi escort protein	810	39

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		melanogaster			
1003	gi10728147	Drosophila melanogaster	CG8356-PA	810	39
1004	gi12652851	Homo sapiens	potassium channel modulatory factor	1987	100
1004	gi4838557	Mus musculus	DEBT-91	1453	96
1004	gi16768790	Drosophila melanogaster	LD03515p	876	63
1005	gi7270532	Arabidopsis thaliana	DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain	173	29
1005	gi16505	Arabidopsis thaliana	RNA polymerase II	173	29
1005	gi16494	Arabidopsis thaliana	DNA-directed RNA polymerase	173	29
1006	gi11875318	Mus musculus	synaptotagmin XIII	2004	89
1006	gi21410154	Mus musculus	synaptotagmin 13	2004	89
1006	gi11119239	Rattus norvegicus	synaptotagmin 13	2000	89
1007	gi3800881	Homo sapiens	RanBP7/importin 7	5447	100
1007	gi11342591	Mus musculus	RanBP7/importin 7	5418	99
1007	gi11544639	Homo sapiens	importin7	5307	100
1008	gi5578958	Homo sapiens	dJ475B7.2 (novel protein)	3770	99
1008	gi18676522	Homo sapiens	FLJ00158 protein	1512	100
1008	gi21595156	Mus musculus	Similar to RIKEN cDNA 5830482G23 gene	1151	71
1009	gi4406393	Bos taurus	differentiation enhancing factor 1	4699	95
1009	gi4063614	Mus musculus	ADP-ribosylation factor-directed GTPase activating protein isoform a	4694	94
1009	gi4063616	Mus musculus	ADP-ribosylation factor-directed GTPase activating protein isoform b	3186	79
1010	gi16411927	Listeria monocytogenes	lmo2439	57	52
1010	gi16415055	Listeria innocua	lin2533	61	57
1010	gi2983786	Aquifex aeolicus	glucose-1-phosphate thymidyltransferase	70	39
1011	gi9280405	Homo sapiens	adlcan	1631	47
1011	gi13872813	Homo sapiens	fibulin-6	502	28
1011	gi3328186	Caenorhabditis elegans	hemicentin precursor	539	27
1012	gi4001698	Sus scrofa	mat-8	67	30
1012	gi2622724	Methanothermobacter thermautotrophicus str. Delta H	conserved protein	82	29
1012	gi498166	Mus musculus	zona-pellucida-binding protein (sp38)	85	27
1013	gi17511816	Homo sapiens	Similar to RIKEN cDNA 1110032O22 gene	1468	99
1013	gi7211438	Homo sapiens	golgin-67	100	30
1013	gi6003208	Human immunodeficiency virus type 1	p17 protein	84	29
1014	gi17511816	Homo sapiens	Similar to RIKEN cDNA 1110032O22 gene	878	100
1014	gi6003208	Human immunodeficiency virus type 1	p17 protein	84	29

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1014	gi21957065	Yersinia pestis KIM	uroporphyrinogen III methylase	90	34
1015	gi2246401	Homo sapiens	centrin	842	100
1015	gi13529248	Homo sapiens	centrin, EF-hand protein, 3 (CDC31 yeast homolog)	839	99
1015	gi2246424	Mus musculus	centrin	832	98
1016	gi17428765	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	530	43
1016	gi15155946	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_C_1725p	379	41
1016	gi15073913	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	372	39
1017	gi17428765	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	381	43
1017	gi15073913	Sinorhizobium meliloti	CONSERVED HYPOTHETICAL PROTEIN	367	48
1017	gi12543118	Corynebacterium glutamicum	RXC01693	265	30
1018	gi6693701	Homo sapiens	melanopsin	2234	91
1018	gi21928729	Homo sapiens	seven transmembrane helix receptor	2190	99
1018	gi6693703	Mus musculus	melanopsin	1735	73
1019	gi439296	Homo sapiens	garp	822	37
1019	gi6572272	Homo sapiens	dJ756G23.1 (novel Leucine Rich Protein)	243	34
1019	gi19344010	Homo sapiens	insulin-like growth factor binding protein, acid labile subunit	293	29
1020	gi15706421	Homo sapiens	middle-chain acyl-CoA synthetase1	1346	99
1020	gi15487302	Homo sapiens	medium-chain acyl-CoA synthetase	1346	99
1020	gi5019275	Bos taurus	xenobiotic/medium-chain fatty acid:CoA ligase form XL-III	1088	78
1021	gi6650766	Homo sapiens	PDZ domain-containing guanine nucleotide exchange factor I	6216	100
1021	gi20386206	Homo sapiens	PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2	5822	98
1021	gi18874700	Homo sapiens	Rap1 guanine nucleotide-exchange factor PDZ-GEF2B	5803	98
1022	gi20386206	Homo sapiens	PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2	5942	100
1022	gi18874700	Homo sapiens	Rap1 guanine nucleotide-exchange factor PDZ-GEF2B	5923	99
1022	gi18874698	Homo sapiens	Rap1 guanine nucleotide-exchange factor PDZ-GEF2A	5923	99
1023	gi13810306	Homo sapiens	transmembrane protein 7	268	37
1023	gi18250724	Mus musculus	transmembrane protein 7	264	37
1023	gi20270907	Oncorhynchus mykiss	VHSV-induced protein-5	243	33
1024	gi21779869	Homo sapiens	IL-17RE	2896	100
1024	gi21779866	Mus musculus	IL-17RE	1394	74
1024	gi21779857	Homo sapiens	IL-17RC	246	29
1025	gi21779869	Homo sapiens	IL-17RE	2928	100
1025	gi21779866	Mus musculus	IL-17RE	1388	75
1025	gi21779857	Homo sapiens	IL-17RC	246	29
1026	gi14150450	Rattus norvegicus	UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T9	1352	93
1026	gi16769916	Drosophila	SD10722p	473	38

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		melanogaster			
1026	gi21627105	Drosophila melanogaster	CG30463-PA	417	38
1027	gi15217067	Homo sapiens	stem cell factor isoform 1	1013	95
1027	gi337934	Homo sapiens	stem cell factor	1013	95
1027	gi1827477	Felis catus	stem cell factor	893	84
1028	gi1377894	Homo sapiens	OB-cadherin-1	1478	64
1028	gi1377895	Homo sapiens	OB-cadherin-2	1478	64
1028	gi506404	Homo sapiens	cadherin-11	1474	63
1029	gi1377894	Homo sapiens	OB-cadherin-1	1628	56
1029	gi1377895	Homo sapiens	OB-cadherin-2	1628	56
1029	gi506404	Homo sapiens	cadherin-11	1623	56
1030	gi1398903	Mus musculus	Ca ²⁺ dependent activator protein for secretion	6314	90
1030	gi577428	Rattus norvegicus	Ca ²⁺ -dependent activator protein; calcium-dependent actin-binding protein	5003	96
1030	gi6980012	Drosophila melanogaster	secretion calcium-dependent activator protein	3540	60
1031	gi217705	Sus scrofa	dipeptidase precursor	781	51
1031	gi2102	Sus scrofa	dipeptidase	781	51
1031	gi8248922	Homo sapiens	renal dipeptidase; RDP	762	50
1032	gi18073362	Homo sapiens	cystine/glutamate transporter	2552	100
1032	gi11493652	Homo sapiens	calcium channel blocker resistance protein CCB1	2552	100
1032	gi13924720	Homo sapiens	cystine/glutamate transporter xCT	2552	100
1033	gi17028348	Homo sapiens	Similar to methylenetetrahydrofolate dehydrogenase (NADP ⁺ dependent), methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase	3748	100
1033	gi20987924	Mus musculus	Similar to DKFZP586G1517 protein	3473	92
1033	gi307178	Homo sapiens	MDMCSF (EC 1.5.1.5; EC 3.5.4.9; EC 6.3.4.3)	2839	62
1034	gi632676	Saccharomyces cerevisiae	Ylr410wp	598	44
1034	gi4070	Saccharomyces cerevisiae	nuf1	120	20
1034	gi312175	Saccharomyces cerevisiae	SPC110/NUF1	120	20
1035	gi11066463	Rattus norvegicus	RhoGEF glutamate transport modulator GTRAP48	5589	80
1035	gi19387126	Mus musculus	guanine nucleotide exchange factor	1794	37
1035	gi7110160	Homo sapiens	guanine nucleotide exchange factor	1792	37
1036	gi2921821	Rattus norvegicus	cytochrome P450 IIE1	68	28
1036	gi8515399	Human respiratory syncytial virus	attachment glycoprotein G	64	29
1036	gi5901834	Drosophila melanogaster	BcDNA.GH09358	95	23
1037	gi17128288	synthetic construct	Primer 1	1689	100
1037	gi20269957	Sus scrofa	phospholipase C delta 4	1469	85
1037	gi21307610	Mus musculus	phospholipase C delta 4	1327	77
1038	gi6978948	Homo sapiens	vaccinia related kinase 3	76	24

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1038	gi349667	Carnobacterium piscicola	carnobacteriocin A	60	41
1038	gi406315	Carnobacterium piscicola	piscicolin 61	60	41
1039	gi4159884	Homo sapiens	similar to mouse olfactory receptor 13; similar to P34984 (PID:g464305)	1597	99
1039	gi9368991	Homo sapiens	dJ1005H11.1 (7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN))	1410	100
1039	gi18480186	Mus musculus	olfactory receptor MOR261-6	1323	81
1040	gi311626	Homo sapiens	thrombospondin-4	4787	99
1040	gi3860231	Mus musculus	thrombospondin-4	4557	93
1040	gi929835	Rattus norvegicus	thrombospondin-4	4547	93
1041	gi14043083	Homo sapiens	sperm associated antigen 9	660	100
1041	gi3116015	Homo sapiens	sperm specific protein	273	98
1041	gi10801148	Mus musculus	JNK/SAPK-associated protein 1	98	41
1042	gi21654741	Homo sapiens	peptide/histidine transporter	1746	98
1042	gi2208839	Rattus norvegicus	peptide/histidine transporter	1469	79
1042	gi16740719	Mus musculus	Similar to peptide transporter 3	1453	83
1043	gi21392228	Drosophila melanogaster	RH61354p	1221	41
1043	gi19353264	Homo sapiens	Similar to dishevelled associated activator of morphogenesis 2	2224	65
1043	gi2947238	Homo sapiens	diaphanous 1	717	32
1044	gi15929979	Homo sapiens	Similar to zinc finger protein 345	2476	100
1044	gi18643896	Homo sapiens	zinc finger protein	1656	53
1044	gi1020145	Homo sapiens	DNA binding protein	1656	53
1045	gi12655913	Homo sapiens	sprouty-4A	386	98
1045	gi4850326	Mus musculus	sprouty-4	323	81
1045	gi5917720	Mus musculus	sprouty 4	323	81
1046	gi4539525	Homo sapiens	NAALADase II protein	3881	100
1046	gi3211746	Sus scrofa	folylpoly-gamma-glutamate carboxypeptidase	2824	70
1046	gi2897946	Homo sapiens	prostate-specific membrane antigen	2787	69
1047	gi5420389	Leishmania major	proteophosphoglycan	139	23
1047	gi915207	Sus scrofa	gastric mucin	123	22
1047	gi13592175	Leishmania major	ppg3	125	23
1048	gi5918167	Homo sapiens	plexin-B1/SEP receptor	2104	54
1048	gi6010211	Homo sapiens	semaphorin receptor	2103	54
1048	gi1655432	Mus musculus	plexin 2	1517	30
1049	gi15990515	Homo sapiens	Similar to RIKEN cDNA 0610020I02 gene	3035	100
1049	gi18380977	Mus musculus	RIKEN cDNA 0610020I02 gene	2792	92
1049	gi2384732	Rattus norvegicus	NAC-1 protein	1269	57
1050	gi15088540	Homo sapiens	sterolin-2	3127	99
1050	gi11692802	Homo sapiens	ABCG8	3123	99
1050	gi15146444	Homo sapiens	sterolin-2	3120	99
1051	gi12652851	Homo sapiens	potassium channel modulatory factor	1987	100

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1051	gi4838557	Mus musculus	DEBT-91	1453	96
1051	gi16768790	Drosophila melanogaster	LD03515p	876	63
1052	gi33730	Homo sapiens	immunoglobulin lambda light chain	716	71
1052	gi33395	Homo sapiens	lambda-chain precursor (AA -20 to 215)	703	70
1052	gi33744	Homo sapiens	immunoglobulin lambda light chain	697	68
1053	gi21388773	Homo sapiens	kringle-containing protein	1552	100
1053	gi21623530	Homo sapiens	kringle-containing transmembrane protein	1238	99
1053	gi21388775	Homo sapiens	kringle-containing protein	1241	100
1054	gi14495324	Homo sapiens	CMRF35A	421	48
1054	gi18490143	Homo sapiens	CMRF35 leukocyte immunoglobulin-like receptor	421	48
1054	gi396170	Homo sapiens	CMRF-35 antigen	421	48
1055	gi4468256	Homo sapiens	MHC class I antigen	1974	100
1055	gi32139	Homo sapiens	HLA-A11E protein precursor (AA -24 to 341)	1912	97
1055	gi487909	Homo sapiens	HLA-A11 antigen A11.1	1912	97
1056	gi21667214	Homo sapiens	bactericidal/permeability-increasing protein-like 3	741	100
1056	gi57732	Rattus rattus	potential ligand-binding protein	215	35
1056	gi11877276	Homo sapiens	dJ726C3.5 (ortholog of potential ligand binding protein RY2G5 (Rat))	176	32
1057	gi21667214	Homo sapiens	bactericidal/permeability-increasing protein-like 3	2226	99
1057	gi57732	Rattus rattus	potential ligand-binding protein	579	32
1057	gi11877276	Homo sapiens	dJ726C3.5 (ortholog of potential ligand binding protein RY2G5 (Rat))	540	31
1058	gi21667214	Homo sapiens	bactericidal/permeability-increasing protein-like 3	1919	99
1058	gi57732	Rattus rattus	potential ligand-binding protein	485	33
1058	gi11877276	Homo sapiens	dJ726C3.5 (ortholog of potential ligand binding protein RY2G5 (Rat))	447	31
1059	gi21667214	Homo sapiens	bactericidal/permeability-increasing protein-like 3	1842	100
1059	gi57732	Rattus rattus	potential ligand-binding protein	485	33
1059	gi11877276	Homo sapiens	dJ726C3.5 (ortholog of potential ligand binding protein RY2G5 (Rat))	447	31
1060	gi23911	Homo sapiens	polypeptide 7B2 precursor	1148	100
1060	gi7718079	Homo sapiens	neuroendocrine protein 7B2	1148	100
1060	gi13529158	Homo sapiens	secretory granule, neuroendocrine protein 1 (7B2 protein)	1131	99
1061	gi18698601	Homo sapiens	Smith-Magenis syndrome chromosome region candidate 7 protein	2325	100
1061	gi15073752	Sinorhizobium meliloti	HYPOTHETICAL TRANSMEMBRANE SIGNAL PEPTIDE PROTEIN	90	29
1061	gi13623063	Streptococcus pyogenes M1 GAS	heat shock protein - cochaperonin	70	32
1062	gi4128041	Homo sapiens	claudin-9 protein	1116	100
1062	gi4325296	Mus musculus	claudin-9	1078	95
1062	gi14286272	Homo sapiens	claudin 6	826	71
1063	gi14286258	Homo sapiens	ribosomal protein L29	432	65
1063	gi1215742	Homo sapiens	HIP	432	65
1063	gi793843	Homo sapiens	ribosomal protein L29	432	65

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1064	gi6601555	Rattus norvegicus	glutamate receptor interacting protein 2	3560	86
1064	gi3639077	Rattus norvegicus	AMPA receptor binding protein	2743	88
1064	gi1890856	Rattus norvegicus	AMPA receptor interacting protein GRIP	1925	59
1065	gi3288852	Homo sapiens	disabled-1	2865	99
1065	gi1771282	Mus musculus	mDab555 protein	2797	96
1065	gi22095317	Gallus gallus	disabled-1	2630	90
1066	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	164	86
1066	gi4336401	Homo sapiens	beta glucuronidase isoform d	127	72
1066	gi4336402	Homo sapiens	beta glucuronidase isoform c	127	72
1067	gi15430703	Homo sapiens	testis specific serine/threonine kinase 2	1858	99
1067	gi2738898	Mus musculus	protein kinase	1686	89
1067	gi15283993	Homo sapiens	testis-specific serine/threonine kinase 1	1230	77
1068	gi13543568	Homo sapiens	prostaglandin D2 synthase (21kD, brain)	977	96
1068	gi12963879	Homo sapiens	prostaglandin D synthase	977	96
1068	gi189772	Homo sapiens	prostaglandin D2 synthase	977	96
1069	gi13279311	Homo sapiens	Similar to RIKEN cDNA 1500017E18 gene	1416	96
1069	gi14336718	Homo sapiens	similar to HAGH	1157	100
1069	gi20988885	Mus musculus	RIKEN cDNA 1500017E18 gene	1151	79
1070	gi13397835	Homo sapiens	annexin A13 isoform b	1795	99
1070	gi757784	Canis familiaris	annexin XIIIb	1621	89
1070	gi21218387	Oryctolagus cuniculus	annexin XIIIb	1589	88
1071	gi21707908	Homo sapiens	solute carrier family 6 (neurotransmitter transporter, GABA), member 1	3129	98
1071	gi31658	Homo sapiens	GABA transporter	3114	98
1071	gi204222	Rattus norvegicus	GABA transporter protein	3097	96
1072	gi7160975	Homo sapiens	voltage-gated sodium channel beta-3 subunit	834	100
1072	gi7161889	Rattus norvegicus	voltage-gated sodium channel beta-3 subunit	823	98
1072	gi14165176	Rattus norvegicus	sodium channel beta 3 subunit	823	98
1074	gi18676470	Homo sapiens	FLJ00132 protein	2515	99
1074	gi21430928	Drosophila melanogaster	SD27341p	324	38
1074	gi20197056	Arabidopsis thaliana	expressed protein	206	29
1075	gi452751	Gallus gallus	Gal beta 1,4 GlcNAc alpha 2,6-sialyltransferase	949	54
1075	gi2295223	unidentified	GALACTOSYLTRANSFERASE-SIALYLTRANSFERASE HYBRID PROTEIN	856	48
1075	gi29434	Homo sapiens	beta-galactoside alpha-2,6-sialyltransferase	856	48
1076	gi13344997	Homo sapiens	Cat Eye Syndrome critical region protein isoform 2	2223	100
1076	gi13344995	Homo sapiens	Cat Eye Syndrome critical region protein isoform 1	2002	99
1076	gi15928451	Mus musculus	Similar to cat eye syndrome chromosome region, candidate 5	1649	76

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1077	gi13344997	Homo sapiens	Cat Eye Syndrome critical region protein isoform 2	1662	96
1077	gi13344995	Homo sapiens	Cat Eye Syndrome critical region protein isoform 1	1662	96
1077	gi15928451	Mus musculus	Similar to cat eye syndrome chromosome region, candidate 5	1294	75
1078	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2714	39
1078	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	2708	39
1078	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	2700	39
1079	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	1300	34
1079	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	1297	35
1079	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	1296	35
1080	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	806	32
1080	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	769	31
1080	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	769	31
1081	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2732	40
1081	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	2726	40
1081	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	2718	39
1082	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	1297	35
1082	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	1296	35
1082	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	1296	35
1083	gi404389	Mus sp.	carboxylesterase; Es-male	2006	66
1083	gi213101	Anas platyrhynchos	thioesterase B	1261	46
1083	gi2058318	Homo sapiens	carboxylesterase	1253	47
1084	gi207286	Rattus norvegicus	TGF-beta masking protein large subunit	8731	89
1084	gi3493176	Mus musculus	latent TGF beta binding protein	8640	88
1084	gi19909128	Homo sapiens	transforming growth factor-beta binding protein-1S	7763	99
1085	gi17985371	Homo sapiens	I3 binding protein	861	100
1085	gi21961229	Homo sapiens	BRI3 binding protein	861	100
1085	gi18466808	Homo sapiens	cervical cancer 1 proto-oncogene-binding protein KG19	853	99
1086	gi222833	Gallus gallus	M-protein	2953	42
1086	gi407097	Homo sapiens	165kD protein	2933	42
1086	gi2950347	Mus musculus	M-protein	2931	42
1087	gi12655165	Homo sapiens	zinc finger protein 256	696	65
1087	gi4894364	Homo sapiens	zinc finger protein 3	696	65
1087	gi21327296	Homo sapiens	zinc finger protein 382	495	46
1088	gi2689441	Homo sapiens	F18547_1	188	37
1088	gi1613848	Homo sapiens	zinc finger protein zfp6	316	49
1088	gi21327296	Homo sapiens	zinc finger protein 382	203	38
1089	gi12655460	Homo sapiens	keratin associated protein 4.12	929	75
1089	gi13278825	Homo sapiens	Similar to RIKEN cDNA 1110054P19 gene	929	75
1089	gi12655464	Homo sapiens	keratin associated protein 4.15	900	83
1090	gi12655460	Homo sapiens	keratin associated protein 4.12	403	85
1090	gi13278825	Homo sapiens	Similar to RIKEN cDNA 1110054P19 gene	403	85
1090	gi12655442	Homo sapiens	keratin associated protein 4.2	397	84
1091	gi12655464	Homo sapiens	keratin associated protein 4.15	1260	100
1091	gi12655452	Homo sapiens	keratin associated protein 4.7	1222	90
1091	gi12655460	Homo sapiens	keratin associated protein 4.12	1156	88

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1092	gi15722084	Homo sapiens	bA304I5.1 (novel lipase)	1991	100
1092	gi21594466	Mus musculus	RIKEN cDNA 4632427C23 gene	1928	87
1092	gi460143	Homo sapiens	lysosomal acid lipase/cholesteryl ester hydrolase	1290	60
1093	gi21594466	Mus musculus	RIKEN cDNA 4632427C23 gene	1957	88
1093	gi15722084	Homo sapiens	bA304I5.1 (novel lipase)	1935	100
1093	gi460143	Homo sapiens	lysosomal acid lipase/cholesteryl ester hydrolase	1290	60
1094	gi8118040	Homo sapiens	orphan G-protein coupled receptor	1804	99
1094	gi8118052	Mus musculus	orphan G-protein coupled receptor	1306	82
1094	gi13177796	Homo sapiens	retinoic acid induced 3	728	45
1095	gi18129609	Homo sapiens	diacylglycerol acyltransferase 2	600	49
1095	gi15099951	Mus musculus	diacylglycerol acyltransferase 2	599	49
1095	gi17426446	Homo sapiens	bA351K23.5 (novel protein)	572	54
1096	gi17225337	Homo sapiens	dendritic lectin	1134	95
1096	gi17224598	Homo sapiens	blood dendritic cell antigen 2 protein	1134	95
1096	gi17225339	Homo sapiens	dendritic lectin b isoform	918	94
1097	gi17225337	Homo sapiens	dendritic lectin	1182	99
1097	gi17224598	Homo sapiens	blood dendritic cell antigen 2 protein	1182	99
1097	gi17225339	Homo sapiens	dendritic lectin b isoform	966	99
1098	gi21929119	Homo sapiens	seven transmembrane helix receptor	1595	100
1098	gi18479834	Mus musculus	olfactory receptor MOR144-1	1223	77
1098	gi18480806	Mus musculus	olfactory receptor MOR143-1	1163	70
1099	gi5911169	Homo sapiens	transmembrane mucin 12	3049	99
1099	gi19526645	Homo sapiens	intestinal membrane mucin MUC17	815	32
1099	gi5911171	Homo sapiens	mucin 11	684	47
1100	gi37198	Homo sapiens	TM1-CEA preprotein	455	34
1100	gi179440	Homo sapiens	biliary glycoprotein I precursor	455	34
1100	gi550031	Homo sapiens	BGPc	455	34
1101	gi6273399	Homo sapiens	melanoma-associated antigen MG50	4733	60
1101	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	4733	60
1101	gi531385	Drosophila melanogaster	peroxidase precursor	2013	39
1102	gi6273399	Homo sapiens	melanoma-associated antigen MG50	4458	60
1102	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	4458	60
1102	gi531385	Drosophila melanogaster	peroxidase precursor	2013	39
1103	gi7264653	Mus musculus	Kiaa0575	2398	61
1103	gi11611734	Homo sapiens	GREB1a	513	46
1103	gi915208	Sus scrofa	gastric mucin	128	30
1104	gi20219008	Chlamydomonas reinhardtii	coiled-coil flagellar protein	682	36
1104	gi16519041	Drosophila melanogaster	occludin-like protein	203	23
1104	gi3549261	Dictyostelium discoideum	interaptin	175	22
1105	gi12654511	Homo sapiens	ATP-dependant interferon response protein 1	693	96
1105	gi17390689	Homo sapiens	ATP-dependant interferon responsive	693	96
1105	gi10862826	Homo sapiens	ADIR1	689	95
1106	gi15215375	Homo sapiens	RNA binding motif protein 12	325	72
1106	gi21666372	Homo sapiens	swan	325	72

174

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1106	gi19070194	Homo sapiens	SWAN	325	72
1107	gi18157547	Mus musculus	pecanex-like 3	3262	97
1107	gi6650377	Mus musculus	pecanex 1	2530	74
1107	gi15076843	Homo sapiens	pecanex-like protein 1	2526	74
1108	gi18157547	Mus musculus	pecanex-like 3	3138	97
1108	gi6650377	Mus musculus	pecanex 1	2409	73
1108	gi15076843	Homo sapiens	pecanex-like protein 1	2405	73
1109	gi7770237	Homo sapiens	PRO2822	233	59
1109	gi21595759	Homo sapiens	similar to HC6	211	71
1109	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	209	67
1110	gi18159337	Pyrobaculum aerophilum	paREP8	77	30
1110	gi1658310	Homo sapiens	leukocyte surface protein	97	26
1110	gi7638235	Mus musculus	immunoglobulin heavy chain variable domain	77	25
1111	gi4263743	Homo sapiens	similar to UNC-93; similar to U89424 (PID:g3642687)	1575	100
1111	gi12043567	Homo sapiens	unc-93 related protein	1571	99
1111	gi17390915	Mus musculus	Similar to unc93 (C.elegans) homolog B	1372	87
1113	gi4153873	Homo sapiens	similar to wee1-like protein kinase; similar to P30291 (PID:g1351419)	2810	100
1113	gi644770	Xenopus laevis	Wee1A kinase	1166	64
1113	gi2827996	Xenopus laevis	wee1 homolog	1166	64
1114	gi6606119	Dothidea inculpta	DNA-dependent RNA polymerase II RPB140	81	32
1114	gi2796053	Mus musculus	T cell receptor beta chain	54	48
1115	gi20372871	Clarkia similis	cytosolic phosphoglucose isomerase	56	28
1116	gi21708029	Homo sapiens	similar to Alu subfamily SQ sequence contamination warning entry	135	70
1116	gi11493409	Homo sapiens	PRO0898	129	59
1116	gi6650818	Homo sapiens	PRO1992	110	70
1117	gi13810898	Rattus norvegicus	inhibin binding protein long isoform	310	37
1117	gi2645890	Homo sapiens	IGSF1	326	40
1117	gi2370143	Homo sapiens	immunoglobulin-like domain-containing 1	326	40
1118	gi13810898	Rattus norvegicus	inhibin binding protein long isoform	310	37
1118	gi2645890	Homo sapiens	IGSF1	312	38
1118	gi2370143	Homo sapiens	immunoglobulin-like domain-containing 1	312	38
1119	gi21707128	Homo sapiens	Ran binding protein 11	5047	99
1119	gi20987296	Mus musculus	Similar to Ran binding protein 11	4898	96
1119	gi17862636	Drosophila melanogaster	LD41918p	1191	38
1120	gi18652832	Homo sapiens	ASPP1 protein	5703	99
1120	gi16197705	Homo sapiens	ASPP2 protein	1556	42
1120	gi1399805	Homo sapiens	Bbp/53BP2	1556	42
1121	gi18448478	Aotus trivirgatus	chorionic gonadotropin beta subunit	47	59
1121	gi5670272	Human herpesvirus 8	K1 glycoprotein	67	38
1121	gi9886851	Human herpesvirus 8	K1 protein	63	36

175

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1122	gi2598461	Homo sapiens	dJ408N23.1 (suppression of tumorigenicity 13 (colon carcinoma) (Hsp70-interacting protein) (Progesterone receptor associated P48 protein))	1887	97
1122	gi904032	Homo sapiens	p48	1869	96
1122	gi21218374	Homo sapiens	FAM10A5	1814	93
1123	gi8927428	Homo sapiens	otoraplin	676	100
1123	gi12619173	Homo sapiens	melanoma inhibitory activity like protein	676	100
1123	gi11323317	Homo sapiens	dJ705D16.2 (Otoraplin)	676	100
1124	gi12034719	Mus musculus	ankyrin-like protein	462	46
1124	gi13469729	Homo sapiens	breast cancer antigen NY-BR-1	448	50
1124	gi21618588	Homo sapiens	testis-specific ankyrin motif containing protein	381	47
1125	gi13469729	Homo sapiens	breast cancer antigen NY-BR-1	364	51
1125	gi12034719	Mus musculus	ankyrin-like protein	379	46
1125	gi21618588	Homo sapiens	testis-specific ankyrin motif containing protein	345	49
1126	gi7770139	Homo sapiens	PRO1722	263	60
1126	gi11493483	Homo sapiens	PRO2550	263	67
1126	gi8572229	Homo sapiens	ubiquitous TPR-motif protein Y isoform	249	61
1127	gi6907090	Oryza sativa (japonica cultivar-group)	Similar to Oryza sativa root-specific RCc3 mRNA. (L27208)	86	35
1127	gi5902450	Cercopithecine herpesvirus 1	glycoprotein G	58	41
1127	gi12750734	Homo sapiens	L-type voltage-dependent calcium channel	56	48
1128	gi16878260	Homo sapiens	Similar to angiotensin II, type I receptor-associated protein	726	100
1128	gi16588454	Homo sapiens	AGTRAP protein	705	95
1128	gi9621816	Homo sapiens	ATRAP	705	95
1129	gi17986216	Homo sapiens	cell recognition molecule CASPR3	1864	98
1129	gi12330704	Mus musculus	cell recognition molecule CASPR4	1376	71
1129	gi21961652	Mus musculus	cell recognition protein CASPR4	1376	71
1130	gi17986216	Homo sapiens	cell recognition molecule CASPR3	6812	99
1130	gi18390059	Homo sapiens	cell recognition protein CASPR4	4754	70
1130	gi21961652	Mus musculus	cell recognition protein CASPR4	4724	68
1131	gi21552969	Mus musculus	Williams-Beuren syndrome critical region gene 17	3100	97
1131	gi10336504	Homo sapiens	UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase	2020	61
1131	gi11041469	Macaca fascicularis	UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase	1913	58
1132	gi13625176	Homo sapiens	thrombospondin	586	46
1132	gi14627121	Homo sapiens	dJ824F16.3 (novel protein similar to mouse thrombospondin type 1 domain protein R-spondin)	544	46
1132	gi4519541	Mus musculus	thrombospondin type 1 domain	511	43
1133	gi5305333	Mus musculus	protein kinase Myak-S	865	50
1133	gi18314319	Mesocricetus auratus	Mx-interacting protein kinase PKM	865	50
1133	gi5815143	Mus musculus	nuclear body associated kinase 2a	865	50
1134	gi14022292	Mesorhizobium loti	cell division protein	45	36

176

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1134	gi180143	Homo sapiens	CD53 glycoprotein	45	53
1134	gi180141	Homo sapiens	cell surface antigen	45	53
1135	gi14571502	Homo sapiens	calcium-promoted Ras inactivator	4174	99
1135	gi2822157	Homo sapiens	similar to GTPase-activating proteins; 35% similar to JC5047 (PID:g2136083)	3961	99
1135	gi4185294	Homo sapiens	rasGAP-activating-like protein	1898	49
1136	gi11527987	Gallus gallus	immunoglobulin-like receptor CHIR-A	97	30
1136	gi432214	Human immunodeficiency virus type 1	envelope glycoprotein gp120	43	39
1136	gi15026993	Homo sapiens	MUC5AC protein	64	38
1137	gi15128103	Mus musculus	nephronectin	2971	87
1137	gi15430248	Mus musculus	nephronectin long isoform	2640	80
1137	gi16040981	Mus musculus	POEM	2374	87
1139	gi7638247	Homo sapiens	mesenchymal stem cell protein DSCD75	595	100
1139	gi17946258	Drosophila melanogaster	RE58349p	165	34
1139	gi21464462	Drosophila melanogaster	RH58440p	158	36
1140	gi21619491	Homo sapiens	similar to expressed sequence AW049604	235	83
1140	gi6572294	Homo sapiens	bA262A13.1 (novel protein)	126	48
1140	gi215692	Bacteriophage P4	gop protein	87	28
1141	gi21619491	Homo sapiens	similar to expressed sequence AW049604	454	82
1141	gi6572294	Homo sapiens	bA262A13.1 (novel protein)	239	48
1141	gi215692	Bacteriophage P4	gop protein	84	33
1142	gi20306274	Homo sapiens	testicular haploid expressed gene	1487	80
1142	gi10443967	Homo sapiens	THEG protein	1487	80
1142	gi7416134	Homo sapiens	testis-specific gene	1487	80
1143	gi21928259	Homo sapiens	seven transmembrane helix receptor	1023	100
1143	gi18480746	Mus musculus	olfactory receptor MOR261-10	864	84
1143	gi18480744	Mus musculus	olfactory receptor MOR261-9	858	82
1144	gi21928655	Homo sapiens	seven transmembrane helix receptor	1458	93
1144	gi18480746	Mus musculus	olfactory receptor MOR261-10	1280	79
1144	gi18480744	Mus musculus	olfactory receptor MOR261-9	1258	78
1145	gi1707674	Streptomyces cinnamomeus	elongation factor G	52	34
1146	gi15779092	Homo sapiens	Similar to syntaxin 18	1295	100
1146	gi7707424	Homo sapiens	syntaxin 18	1295	100
1146	gi18203931	Mus musculus	Similar to syntaxin 18	873	90
1147	gi14573319	Homo sapiens	interleukin-1 HY2	812	99
1147	gi18025344	Homo sapiens	interleukin-1 receptor antagonist-like FIL1 theta	809	99
1147	gi19068192	Mus musculus	IL-1F10	662	82
1148	gi4103158	Mus musculus	hair keratin acidic 5; Ha5 keratin	1116	72
1148	gi3724107	Homo sapiens	keratin, type I	1114	72
1148	gi1668744	Homo sapiens	HHa5 hair keratin type I intermediate filament	1114	72
1149	gi19353375	Mus musculus	RIKEN cDNA 1110031I02 gene	1417	84
1149	gi6166378	Mus musculus	growth suppressor 1L	141	30
1149	gi15929776	Homo sapiens	growth suppressor 1	137	41
1150	gi13623421	Homo sapiens	Similar to RIKEN cDNA 5730589L02 gene	1336	90

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1150	gi19484086	Mus musculus	RIKEN cDNA 5730589L02 gene	1287	86
1150	gi1699265	Homo sapiens	malignant cell expression-enhanced gene/tumor progression-enhanced gene	392	57
1151	gi15419605	Canis familiaris	masticatory epithelia keratin 2p	1204	55
1151	gi14595019	Homo sapiens	keratin 6 irs	1175	54
1151	gi6092075	Mus musculus	type II cytokeratin	1116	51
1152	gi11066090	Homo sapiens	matrix metalloprotease MMP-27	1382	96
1152	gi12006364	Tupaia belangeri	matrix metalloproteinase-27	1121	80
1152	gi3511149	Gallus gallus	matrix metalloproteinase	663	57
1153	gi11066090	Homo sapiens	matrix metalloprotease MMP-27	1382	96
1153	gi12006364	Tupaia belangeri	matrix metalloproteinase-27	1121	80
1153	gi3511149	Gallus gallus	matrix metalloproteinase	663	57
1154	gi6689894	Homo sapiens	Suppressor of Fused	2599	100
1154	gi5739507	Homo sapiens	suppressor of fused	2594	99
1154	gi4468628	Mus musculus	Su(fu) protein	2541	97
1155	gi21667212	Homo sapiens	bactericidal/permeability-increasing protein-like 2	2600	100
1155	gi20387085	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein)-1	690	31
1155	gi20387087	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein) like-2	685	30
1156	gi11229139	Homo sapiens	bB152O15.3 (SRY (sex determining region Y)-box 18)	2066	100
1156	gi12082687	Homo sapiens	Sry-related HMG-box protein	2066	100
1156	gi8894593	Homo sapiens	SOX18 protein	2066	100
1157	gi19526647	Homo sapiens	oxidored-nitro domain-containing protein	837	85
1157	gi7303522	Drosophila melanogaster	CG13178-PA	172	31
1157	gi16304788	Mus musculus	bendless-like ubiquitin conjugating enzyme	83	28
1158	gi19526647	Homo sapiens	oxidored-nitro domain-containing protein	837	85
1158	gi7303522	Drosophila melanogaster	CG13178-PA	172	31
1158	gi16304788	Mus musculus	bendless-like ubiquitin conjugating enzyme	83	28
1159	gi1794221	Mus musculus	DNA ligase III-beta	2987	89
1159	gi1794223	Mus musculus	DNA ligase III-alpha	2987	89
1159	gi19550955	Homo sapiens	ligase III, DNA, ATP-dependent	2875	100
1160	gi15667919	Homo sapiens	SERPINB12	1678	99
1160	gi12597188	Homo sapiens	squamous cell carcinoma antigen 2	749	48
1160	gi1235617	Homo sapiens	squamous cell carcinoma antigen	749	48
1161	gi15141587	Eulemur rubriventer	olfactory receptor	67	34
1161	gi21739229	Oryza sativa	OSJNBa0072F16.8	67	43
1161	gi21629328	Leishmania major	L3561.8	65	37
1162	gi2589190	Homo sapiens	skin-specific protein	68	39
1162	gi38232	Pan troglodytes	immunoglobulin alpha heavy chain	61	39
1162	gi14021730	Mesorhizobium loti	c-type cytochrome biogenesis protein	68	31

178

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1163	gi7228149	Mus musculus	ATFa-associated factor	354	50
1163	gi7303705	Drosophila melanogaster	CG12340-PA	193	24
1163	gi5052666	Drosophila melanogaster	BcDNA.LD26050	193	24
1164	gi20901968	Caenorhabditis elegans	C. elegans RPL-36 protein (corresponding sequence F37C12.4)	71	34
1165	gi5911451	Drosophila nannoptera	cytochrome oxidase III	43	41
1165	gi13276253	Homo sapiens	T-cell receptor beta chain VJ region	56	34
1165	gi3928896	Homo sapiens	SH2 domain protein 1A isoform C	55	38
1166	gi20381326	Homo sapiens	Similar to caspase 8, apoptosis-related cysteine protease	263	100
1166	gi14211398	Homo sapiens	caspase-8L	263	100
1166	gi19401524	Homo sapiens	procaspase-8	223	95
1167	gi10440448	Homo sapiens	FLJ00060 protein	1204	98
1167	gi3983420	Homo sapiens	KIR3DL1-like natural killer cell receptor	693	47
1167	gi13560453	Homo sapiens	killer cell immunoglobulin-like receptor 3DL1	693	47
1168	gi1799570	Rattus norvegicus	TIP120	4573	99
1168	gi7688703	Homo sapiens	TIP120 protein	4573	99
1168	gi5811583	Rattus norvegicus	TIP120-family protein TIP120B	2735	57
1169	gi13016701	Homo sapiens	activating coreceptor NKp80	1226	100
1169	gi7188567	Homo sapiens	lectin-like receptor F1	1226	100
1169	gi22449867	Macaca fascicularis	NKp80 NK receptor	1122	90
1170	gi14027275	Mesorhizobium loti	nodulation protein nodG, 3-oxoacyl-(acyl carrier protein) reductase	70	27
1170	gi1531618	Rhizobium sp. N33	NodG	68	26
1170	gi6899062	Ureaplasma urealyticum	seryl-tRNA synthetase	70	31
1171	gi3021409	Homo sapiens	transducin (beta) like 1 protein	3057	100
1171	gi13161069	Homo sapiens	transducin beta-like 1	2548	91
1171	gi12642596	Homo sapiens	nuclear receptor co-repressor/HDAC3 complex subunit TBLR1	2431	86
1172	gi13623421	Homo sapiens	Similar to RIKEN cDNA 5730589L02 gene, clone MGC:13124 IMAGE:4110925, mRNA, complete cds.	380	69
1172	gi12803383	Homo sapiens	clone MGC:2099 IMAGE:3051525, mRNA, complete cds.	376	68
1172	gi13111983	Homo sapiens	clone MGC:4221 IMAGE:2958347, mRNA, complete cds.	376	68
1173	gi13623421	Homo sapiens	Similar to RIKEN cDNA 5730589L02 gene, clone MGC:13124 IMAGE:4110925, mRNA, complete cds.	380	69
1173	gi12803383	Homo sapiens	clone MGC:2099 IMAGE:3051525, mRNA, complete cds.	376	68
1173	gi13111983	Homo sapiens	clone MGC:4221 IMAGE:2958347, mRNA, complete cds.	376	68
1174	gi13623421	Homo sapiens	Similar to RIKEN cDNA 5730589L02 gene	1830	99
1174	gi19484086	Mus musculus	RIKEN cDNA 5730589L02 gene	1802	95

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1174	gi1699265	Homo sapiens	malignant cell expression-enhanced gene/tumor progression-enhanced gene	930	81
1175	gi13182755	Homo sapiens	HPHRP	1210	100
1175	gi15929309	Homo sapiens	phosphotriesterase related	1210	100
1175	gi881499	Mus musculus	parathion hydrolase (phosphotriesterase)-related protein	1069	86
1176	gi552075	Chironomus tentans	giant secretory protein	71	28
1176	gi15419013	Toxoplasma gondii	subtilisin-like protein	71	34
1176	gi156534	Chironomus tentans	giant secretory protein (gsp)	66	28
1177	gi5458910	Pyrococcus abyssi	FLAGELLA-RELATED PROTEIN C	103	24
1177	gi487272	Enterococcus hirae	Na ⁺ -ATPase subunit F	90	31
1177	gi9229886	Ciona intestinalis	ezrin/radixin/moesin (ERM)-like protein	111	27
1178	gi21554060	Arabidopsis thaliana	phytoeyanin	44	43
1178	gi205640	Rattus norvegicus	acetylcholine receptor alpha subunit	53	44
1178	gi4028904	Rattus norvegicus	nicotinic acetylcholine receptor alpha 4 subunit	53	44
1179	gi18375961	Neurospora crassa	related to ARCA protein	53	44
1179	gi2935025	Rhodococcus opacus	protocatechuate dioxygenase alpha subunit	38	38
1179	gi13421646	Caulobacter crescentus CB15	spoU rRNA methylase family protein	39	40
1180	gi14348558	Homo sapiens	cDNA encoding protease domain of endotheliase 1	82	38
1180	gi1245184	Danio rerio	Zg01	66	33
1180	gi6137097	Homo sapiens	serine protease DESC1	82	38
1181	gi19528151	Drosophila melanogaster	AT26759p	59	35
1181	gi16768554	Drosophila melanogaster	GM08606p	59	35
1181	gi7291750	Drosophila melanogaster	CG4065-PA	59	35
1182	gi13377880	Cricetulus longicaudatus	arginine N-methyltransferase p82 isoform	3253	85
1182	gi13377882	Cricetulus longicaudatus	arginine N-methyltransferase p77 isoform	3253	85
1182	gi21626587	Drosophila melanogaster	CG9882-PA	1213	36
1183	gi191185	Cricetulus griseus	phosphatidylserine decarboxylase	1130	88
1183	gi5921491	Homo sapiens	dJ858B16.2 (phosphatidylserine decarboxylase (PSSC, EC 4.1.1.65))	1220	96
1183	gi16306618	Homo sapiens	phosphatidylserine decarboxylase	1220	96
1184	gi11907580	Mus musculus	TSC22-related inducible leucine zipper 3c	894	87
1184	gi5231131	Homo sapiens	TSC-22 related protein	460	98

180

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1184	gi5919161	Homo sapiens	TSC-22-like Protein	460	98
1185	gi13874437	Homo sapiens	cerebral protein-11	1461	68
1185	gi15292367	Drosophila melanogaster	LD47668p	510	41
1185	gi2443444	Homo sapiens	TEX28	310	40
1186	gi13543940	Homo sapiens	Similar to RIKEN cDNA 2610017G09 gene	2568	99
1186	gi18204520	Mus musculus	RIKEN cDNA 2610017G09 gene	2381	91
1186	gi16923351	Homo sapiens	RbBP-35	1434	99
1187	gi18676660	Homo sapiens	FLJ00229 protein	931	91
1187	gi5824711	Caenorhabditis elegans	similar to 7TM chemoreceptor (srd-family)	80	20
1187	gi8825622	Rattus norvegicus	T cell receptor	68	36
1188	gi17865311	Homo sapiens	dipeptidyl peptidase-like protein 9	4646	100
1188	gi11095188	Homo sapiens	dipeptidyl peptidase 8	2876	60
1188	gi21265133	Homo sapiens	Similar to dipeptidylpeptidase 8	2217	58
1189	gi17865311	Homo sapiens	dipeptidyl peptidase-like protein 9	4069	100
1189	gi11095188	Homo sapiens	dipeptidyl peptidase 8	2454	59
1189	gi21265133	Homo sapiens	Similar to dipeptidylpeptidase 8	2455	56
1190	gi17865311	Homo sapiens	dipeptidyl peptidase-like protein 9	4542	98
1190	gi11095188	Homo sapiens	dipeptidyl peptidase 8	2810	60
1190	gi21265133	Homo sapiens	Similar to dipeptidylpeptidase 8	2151	57
1191	gi337508	Homo sapiens	ribosomal protein	554	99
1191	gi57724	Rattus rattus	ribosomal protein S25	554	99
1191	gi12805251	Mus musculus	ribosomal protein S25	554	99
1192	gi208176	synthetic construct	D2-T antigen	61	40
1193	gi7328583	Drosophila melanogaster	mechanosensory transduction channel NOMPC	851	28
1193	gi7385113	Bos taurus	ankyrin 1	777	30
1193	gi11065673	Caenorhabditis elegans	Y71A12B.4	778	28
1194	gi7672669	Homo sapiens	serine protease Htra2	1890	100
1194	gi12652695	Homo sapiens	HtrA-like serine protease	1890	100
1194	gi5870865	Homo sapiens	serine protease	1890	100
1195	gi349449	Homo sapiens	A3 adenosine receptor	904	100
1195	gi13559064	Homo sapiens	bA552M11.6 (adenosine A3 receptor)	904	100
1195	gi20988265	Homo sapiens	adenosine A3 receptor	904	100
1196	gi21645219	Drosophila melanogaster	CG15671-PA	299	37
1196	gi9864185	Drosophila melanogaster	Crossveinless 2	299	37
1196	gi7768636	Xenopus laevis	Kielin	276	34
1197	gi18480772	Mus musculus	olfactory receptor MOR101-2	1415	84
1197	gi18479346	Mus musculus	olfactory receptor MOR101-1	1334	82
1197	gi3769616	Rattus norvegicus	olfactory receptor	973	86
1198	gi498768	Serratia marcescens	Deoxyadenosyl-methyltransferase	339	51
1198	gi10799034	Vibrio cholerae	DNA adenine methylase	332	54
1198	gi10799036	Yersinia pseudotuberculosis	DNA adenine methylase	331	52

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1199	gi16974751	Gallus gallus	CALII	338	37
1199	gi666121	Xenopus laevis	cpl-1	293	33
1199	gi1213589	Xenopus laevis	Prostaglandin D Synthase	292	33
1200	gi22296200	Thermosynecho coccus elongatus BP-1	asparaginyl-tRNA synthetase	1057	46
1200	gi17132791	Nostoc sp. PCC 7120	asparaginyl-tRNA synthetase	1027	46
1200	gi19713460	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	Asparaginyl-tRNA synthetase	1013	43
1201	gi18088970	Homo sapiens	Similar to RIKEN cDNA 4933400E14 gene	1263	99
1201	gi20067381	Homo sapiens	ALMS1 protein	249	41
1201	gi21552774	Mus musculus	Almstrom syndrome 1 protein	219	38
1202	gi347134	Homo sapiens	succinate dehydrogenase flavoprotein subunit	495	92
1202	gi12655061	Homo sapiens	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	495	92
1202	gi506338	Homo sapiens	flavoprotein subunit of complex II	495	92
1203	gi18490322	Homo sapiens	Similar to RIKEN cDNA 6330404M18 gene	2241	99
1203	gi21928186	Mus musculus	GPI-gamma 4; GPIgamma4	1471	61
1203	gi17946082	Drosophila melanogaster	RE54096p	688	47
1204	gi9957165	Homo sapiens	alphaCP-3	1722	100
1204	gi9957161	Mus musculus	alphaCP-3	1708	99
1204	gi15082311	Homo sapiens	Similar to poly(rC)-binding protein 3	840	99
1205	gi14574118	Caenorhabditis elegans	C. elegans DPY-19 protein (corresponding sequence F22B7.10)	239	31
1205	gi12328595	Heterodoxus macropus	NADH dehydrogenase subunit 2	79	29
1205	gi18378695	Bufo maculatus	NADH dehydrogenase subunit 2	75	24
1206	gi189760	Homo sapiens	pyruvate dehydrogenase beta-subunit	1710	96
1206	gi189762	Homo sapiens	pyruvate dehydrogenase E1-beta subunit	1710	96
1206	gi190792	Homo sapiens	pyruvate dehydrogenase E1-beta subunit precursor	1710	96
1207	gi688292	Homo sapiens	calmitine; calsequestrine	2029	100
1207	gi2618621	Mus musculus	skeletal muscle calsequestrin	1938	94
1207	gi164842	Oryctolagus cuniculus	calsequestrin	1908	94
1208	gi22295775	Thermosynecho coccus elongatus BP-1	periplasmic sugar-binding protein of sugar ABC transporter	65	35
1208	gi2622963	Methanothermobacter thermautotrophicus str. Delta H	conserved protein	59	30
1208	gi18377999	Drysdalia coronata	NADH dehydrogenase subunit 1	61	34
1209	gi11034760	Homo sapiens	NIBAN	3692	99
1209	gi10432376	Homo sapiens	bG56G5.1 (novel protein)	3334	99
1209	gi11022733	Mus musculus	Niban	2320	67

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1210	gi2982508	Homo sapiens	TCR beta chain	1292	93
1210	gi3002925	Homo sapiens	T cell receptor beta chain	1281	93
1210	gi36733	Homo sapiens	T cell antigen receptor beta chain	1028	75
1211	gi12006041	Homo sapiens	AD038	761	98
1211	gi14189960	Homo sapiens	PRO0764	141	53
1211	gi19072857	Homo sapiens	lung squamous cell cancer related protein LSCC-3	129	60
1213	gi2995719	Homo sapiens	protocadherin 43	4792	100
1213	gi20072790	Homo sapiens	protocadherin gamma subfamily C, 3	4777	99
1213	gi5456977	Homo sapiens	protocadherin gamma C3	4777	99
1214	gi337487	Homo sapiens	Ro ribonucleoprotein autoantigen (Ro/SS-A) precursor	1747	99
1214	gi179882	Homo sapiens	calreticulin	1747	99
1214	gi22203354	Cricetulus griseus	calreticulin	1687	95
1215	gi200964	Mus musculus	serine 2 ultra high sulfur protein	319	52
1215	gi3228237	Homo sapiens	ultra high sulfur keratin	281	49
1215	gi200962	Mus musculus	serine 1 ultra high sulfur protein	281	50
1216	gi13940422	Macaca sylvanus	ATPase subunit 8	56	31
1217	gi5917716	Gallus gallus	sprouty 2	60	45
1217	gi14275701	Influenza virus	matrix protein 2	62	32
1217	gi2738577	Homo sapiens	connexin46.6	54	50
1218	gi17223709	Homo sapiens	selenoprotein SelM	235	100
1218	gi17223711	Mus musculus	selenoprotein SelM	188	78
1218	gi7380925	Bos taurus	Fc gamma receptor III	73	45
1219	gi15025778	Clostridium acetobutylicum	Predicted membrane protein	50	36
1219	gi13752743	Serratia marcescens	TrpG	65	51
1219	gi20906991	Methanosarcina mazei Goel	Cation transporter	62	29
1220	gi535358	Neisseria gonorrhoeae	Opa15063G	60	50
1220	gi1480793	Neisseria meningitidis	OpaI1	58	47
1221	gi992950	Homo sapiens	OPN-c	1426	98
1221	gi189151	Homo sapiens	nephropontin precursor	1377	90
1221	gi1001963	Homo sapiens	osteopontin	1377	90
1223	gi18088363	Homo sapiens	advanced glycosylation end product-specific receptor	2004	99
1223	gi1841550	Homo sapiens	receptor for advanced glycosylation end products	2004	99
1223	gi6691626	Homo sapiens	advanced glycation endproducts receptor	2004	99
1224	gi3157464	Thermus sp. A4	integral membrane protein	77	38
1224	gi8778370	Arabidopsis thaliana	F15O4.23	65	37
1224	gi15156782	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_C_3106p	59	34
1225	gi37231	Homo sapiens	DNA topoisomerase II	8061	99
1225	gi3869382	Homo sapiens	DNA topoisomerase II beta	8048	99
1225	gi790988	Cricetulus longicaudatus	DNA topoisomerase (ATP-hydrolysing)	7892	97

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1226	gi10041309	Homo sapiens	hFATP1	3336	98
1226	gi1881713	Rattus norvegicus	fatty acid transport protein	3031	87
1226	gi10041307	Rattus sp.	rFATP	3031	87
1227	gi3309176	Mus musculus	COP9 complex subunit 7b	796	94
1227	gi15215085	Mus musculus	Similar to COP9 (constitutive photomorphogenic), subunit 7b (Arabidopsis)	793	93
1227	gi19909525	Homo sapiens	DERP10 (dermal papilla derived protein 10)	467	56
1228	gi6942096	Mus musculus	CBLN3	938	93
1228	gi180251	Homo sapiens	precerebellin	551	58
1228	gi5702371	Mus musculus	precerebellin-1	544	57
1229	gi17861952	Drosophila melanogaster	LD01947p	1384	50
1229	gi6850946	Homo sapiens	dJ322I12.1 (novel protein similar to C. elegans C05C8.6 (Tr:016313))	336	100
1229	gi21411108	Mus musculus	Similar to BTB domain protein BDPL	211	32
1230	gi8132557	Drosophila melanogaster	ankyrin 2	729	30
1230	gi710551	Mus musculus	ankyrin 3	734	29
1230	gi1841966	Rattus norvegicus	ankyrin	700	30
1231	gi21667212	Homo sapiens	bactericidal/permeability-increasing protein-like 2	2384	98
1231	gi20387085	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein)-1	672	31
1231	gi20387087	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein) like-2	667	30
1232	gi21667212	Homo sapiens	bactericidal/permeability-increasing protein-like 2	2389	99
1232	gi20387085	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein)-1	664	31
1232	gi20387087	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein) like-2	659	30
1233	gi21667212	Homo sapiens	bactericidal/permeability-increasing protein-like 2	2595	99
1233	gi20387085	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein)-1	698	31
1233	gi20387087	Oncorhynchus mykiss	LBP (LPS binding protein)/BPI (bactericidal/permeability-increasing protein) like-2	693	30
1234	gi19569876	Dictyostelium discoideum	SIMILAR TO HYPOTHETICAL 26.2 KD PROTEIN	247	26
1234	gi2191168	Arabidopsis thaliana	contains similarity to myosin heavy chain	187	27
1234	gi603379	Saccharomyces cerevisiae	Yer139cp	145	28
1235	gi11493528	Homo sapiens	PRO1953	671	100
1235	gi19912632	Eulemur	MHC class II antigen	56	33

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		rubriventer			
1235	gi19912630	Eulemur macaco macaco	MHC class II antigen	55	33
1236	gi17065951	Ostertagia ostertagi	collagen	70	35
1236	gi158077	Drosophila robusta	period protein	69	38
1236	gi497417	Glycine max	dehydrin-like protein	81	27
1237	gi3068592	Mus musculus	punc	2396	94
1237	gi19570398	Homo sapiens	hDDM36	890	41
1237	gi11862941	Mus musculus	DDM36E	892	41
1238	gi12667401	Homo sapiens	NUF2R	2347	99
1238	gi14317902	Homo sapiens	kinetochore protein Nuf2	2347	99
1238	gi12667403	Mus musculus	NUF2R	1754	73
1239	gi2494126	Arabidopsis thaliana	Contains similarity to Chlamydia outer membrane protein (gb X53512).	94	23
1239	gi19887475	Methanopyrus kandleri AV19	Uncharacterized protein conserved in archaea	68	34
1239	gi21646173	Chlorobium tepidum TLS	ribosomal protein S20	67	29
1240	gi21634825	Homo sapiens	semaphorin 6D isoform 4	5658	98
1240	gi21634823	Homo sapiens	semaphorin 6D isoform 3	3106	96
1240	gi21634827	Homo sapiens	semaphorin 6D isoform 1	3106	99
1241	gi9949555	Pseudomonas aeruginosa	probable pyruvate dehydrogenase E1 component, alpha subunit	71	35
1241	gi48708	Mycobacterium tuberculosis	ORFa1 (AA 1 - 74)	58	37
1241	gi307352	Homo sapiens	prothymosin alpha	54	35
1242	gi9106331	Xylella fastidiosa 9a5c	3-dehydroquinate synthase	43	34
1242	gi13700302	Staphylococcus aureus subsp. aureus N315	xanthine phosphoribosyltransferase	45	35
1242	gi21203529	Staphylococcus aureus subsp. aureus MW2	xanthine phosphoribosyltransferase	45	35
1243	gi21671105	Homo sapiens	RAD52B	1134	100
1243	gi20070921	Mus musculus	RIKEN cDNA 2410008M22 gene	829	74
1243	gi21594785	Homo sapiens	Similar to RIKEN cDNA 2410008M22 gene	572	97
1244	gi6013381	Rattus norvegicus	TM6P1	147	47
1244	gi19353944	Mus musculus	RIKEN cDNA 2610318G18 gene	127	31
1244	gi20270909	Oncorhynchus mykiss	VHSV-induced protein-6	118	31
1245	gi6013381	Rattus norvegicus	TM6P1	272	36
1245	gi21428644	Drosophila melanogaster	LP10820p	256	42
1245	gi20270909	Oncorhynchus mykiss	VHSV-induced protein-6	190	29
1246	gi11993700	Homo sapiens	melastatin 2	1194	100
1246	gi3243075	Homo sapiens	melastatin 1	1057	83
1246	gi3047242	Mus musculus	melastatin	1050	83
1247	gi18044366	Homo sapiens	Similar to MEGF10 protein	3468	99

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1247	gi17386053	Mus musculus	Jedi protein	2280	51
1247	gi18252658	Mus musculus	Jedi-736 protein	2280	51
1248	gi20987880	Mus musculus	Similar to PTH-responsive osteosarcoma B1 protein	3586	87
1248	gi4588087	Homo sapiens	PTH-responsive osteosarcoma B1 protein	2264	92
1248	gi21595711	Homo sapiens	Similar to PTH-responsive osteosarcoma B1 protein	1546	100
1249	gi19913471	Homo sapiens	similar to dJ84N20.1.1 (novel protein, isoform 1)	1265	99
1249	gi13591434	Homo sapiens	dJ84N20.1.2 (novel protein, isoform 2)	1160	100
1249	gi13591435	Homo sapiens	dJ84N20.1.1 (novel protein, isoform 1)	976	99
1250	gi16605581	Homo sapiens	H-rev107-like protein 5	1451	100
1250	gi21707989	Homo sapiens	Similar to H-rev107-like protein 5	1376	96
1250	gi6048565	Homo sapiens	retinoid inducible gene 1	382	54
1251	gi21263094	Rattus norvegicus	tramdorin 1	1667	81
1251	gi21263092	Mus musculus	tramdorin 1	1664	82
1251	gi21908026	Mus musculus	proton/amino acid transporter 2	1664	82
1252	gi14571904	Rattus norvegicus	lysosomal amino acid transporter 1	1690	87
1252	gi21908024	Mus musculus	proton/amino acid transporter 1	1685	87
1252	gi21263092	Mus musculus	tramdorin 1	1294	66
1253	gi21595630	Homo sapiens	Similar to forkhead box L2	75	44
1253	gi10580569	Halobacterium sp. NRC-1	trans lesion repair; YqjH	69	51
1253	gi557673	Sus scrofa	BM88 antigen	72	41
1254	gi1669500	Mus musculus	fibroblast growth factor homologous factor 1	917	90
1254	gi1563885	Homo sapiens	fibroblast growth factor homologous factor 1	917	90
1254	gi14317951	Rattus norvegicus	fibroblast growth factor homologous factor 1B	916	98
1255	gi13529143	Homo sapiens	Similar to RIKEN cDNA 1700010H15 gene	779	100
1255	gi19263005	Ciona intestinalis	leucine-rich repeat dynein light chain	759	75
1255	gi2760161	Anthocidaris crassispina	outer arm dynein light chain 2	656	68
1256	gi12666529	Mus musculus	b,b-carotene-9',10'-dioxygenase	2356	80
1256	gi4001821	Ambystoma tigrinum	RPE65 protein; retinal pigment epithelium 65-protein	1125	44
1256	gi11990268	Mus musculus	beta,beta-carotene 15,15'-dioxygenase	1110	42
1257	gi12666529	Mus musculus	b,b-carotene-9',10'-dioxygenase	2305	81
1257	gi4001821	Ambystoma tigrinum	RPE65 protein; retinal pigment epithelium 65-protein	1122	44
1257	gi11990268	Mus musculus	beta,beta-carotene 15,15'-dioxygenase	1113	42
1258	gi18490501	Mus musculus	RIKEN cDNA 2010002A20 gene	868	76
1258	gi61	Bos taurus	calmodulin-independent adenylate cyclase	166	29
1258	gi15559697	Homo sapiens	Similar to neural cell adhesion molecule 1	165	29
1259	gi21748488	Homo sapiens	FLJ00277 protein	50	52
1259	gi2331293	Mus musculus	preprocortistatin	73	40
1259	gi1335910	Rattus norvegicus	preprocortistatin	58	36

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1260	gi1079734	Mus musculus	citron	1291	94
1260	gi3599509	Mus musculus	rho/rac-interacting citron kinase	1286	94
1260	gi2745840	Rattus norvegicus	postsynaptic density protein; citron	1262	93
1261	gi14715029	Mus musculus	serine (or cysteine) proteinase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2	407	39
1261	gi551065	Mus musculus	protease-nexin 1	406	38
1261	gi412157	Homo sapiens	glia-derived neurite-promoting factor (GdNPF)	397	38
1262	gi4323581	Homo sapiens	senescence-associated epithelial membrane protein	223	97
1262	gi15214678	Homo sapiens	claudin 1	223	97
1262	gi7381083	Homo sapiens	claudin-1	223	97
1263	gi21634445	Homo sapiens	GTP-binding protein Sara	449	57
1263	gi13542685	Mus musculus	SAR1 protein	446	54
1263	gi8926205	Homo sapiens	SAR1	445	54
1264	gi11558264	Homo sapiens	sphingosine-1-phosphatase	697	37
1264	gi13447199	Homo sapiens	sphingosine-1-phosphate phosphatase	683	37
1264	gi9623190	Mus musculus	sphingosine-1-phosphate phosphohydrolase	691	38
1265	gi14	Bos taurus	BoWC1.1	1026	37
1265	gi5107945	Homo sapiens	CD163	1093	40
1265	gi312142	Homo sapiens	M130 antigen	1093	40
1266	gi14	Bos taurus	BoWC1.1	1026	37
1266	gi5107945	Homo sapiens	CD163	1093	40
1266	gi312142	Homo sapiens	M130 antigen	1093	40
1267	gi18873700	Necator americanus	NADH dehydrogenase subunit 2	69	32
1267	gi20338417	Gallus gallus	potassium channel subunit	57	31
1267	gi396416	Escherichia coli	similar to Neurospora crassa phosphate-repressible phosphate permease	72	42
1268	gi21619491	Homo sapiens	similar to expressed sequence AW049604	778	100
1268	gi6572294	Homo sapiens	bA262A13.1 (novel protein)	251	49
1268	gi161662	Tribolium castaneum	zinc finger protein	60	26
1269	gi21591552	Haemophilus influenzae biotype aegyptius	cell filamentation-like protein	55	31
1269	gi1762771	Pleurodeles waltl	homeodomain-containing protein	66	35
1269	gi19528253	Drosophila melanogaster	GH13327p	53	41
1270	gi18033185	Danio rerio	UNC45-related protein	3103	73
1270	gi12248757	Homo sapiens	SMAP-1	2393	57
1270	gi12248771	Homo sapiens	SMAP-1b	2393	57
1271	gi21064657	Drosophila melanogaster	RH01479p	185	39
1271	gi7304173	Drosophila melanogaster	CG1577-PA	185	39
1271	gi20150011	Pseudomonas fluorescens	MmpIV	89	36
1272	gi9366656	Trypanosoma	probable similar to ring-h2 finger protein	76	55

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		brucei	rhala.		
1272	gi6714271	Arabidopsis thaliana	F6N18.7	59	36
1272	gi10440424	Homo sapiens	FLJ00047 protein	74	50
1273	gi15823642	Homo sapiens	ALS2CR7	2038	100
1273	gi2645810	Mus musculus	Pftaire-1	1195	68
1273	gi2392814	Mus musculus	PFTAIRE kinase	1190	67
1274	gi2407911	Homo sapiens	differentially expressed in Fanconi anemia	714	96
1274	gi21595389	Homo sapiens	similar to FYVE finger-containing phosphoinositide kinase (1-phosphatidylinositol-4-phosphate kinase) (PIP5K) (PtdIns(4)P-5-kinase) (p235)	89	27
1274	gi330134	human herpesvirus 1	latency-related protein 1	87	46
1275	gi21908028	Homo sapiens	a disintegrin and metalloprotease domain 33	4205	97
1275	gi18147612	Homo sapiens	metalloprotease disintegrin	4204	97
1275	gi13157560	Homo sapiens	dJ964F7.1 (novel disintegrin and reprotysin metalloproteinase family protein)	3916	97
1276	gi530876	Chlamydomonas reinhardtii	amino acid feature: Rod protein domain, aa 266 .. 468; amino acid feature: globular protein domain, aa 32 .. 265	138	35
1276	gi141852	Actinomyces viscosus	sialidase	137	30
1276	gi13926258	Arabidopsis thaliana	AT5g10430/F12B17_220	110	34
1277	gi15291913	Drosophila melanogaster	LD31582p	201	36
1277	gi16648042	Drosophila melanogaster	GH07105p	131	39
1277	gi16416111	Neurospora crassa	related to suppressor protein SPT23	129	43
1278	gi544755	Oryctolagus cuniculus	aminopeptidase N; APN	1016	38
1278	gi525287	Sus scrofa	aminopeptidase N.	1012	39
1278	gi205109	Rattus norvegicus	kidney Zn-peptidase precursor	1004	39
1279	gi13559063	Homo sapiens	bA552M11.5 (novel protein)	747	100
1279	gi9963863	Homo sapiens	AD026	738	98
1279	gi19263987	Homo sapiens	similar to CMRF35 ANTIGEN PRECURSOR	131	32
1280	gi2773306	Equus caballus	type II collagen	69	31
1280	gi3687594	Canis familiaris	type IIB procollagen	69	31
1280	gi8918871	YccA of plasmid ColIb-P9] [Plasmid F	96 pct identical to gp:AB021078_30	64	26
1281	gi9927307	Mus musculus	junctophilin type 3	59	42
1281	gi5881591	Gallus gallus	homeodomain protein	78	38
1281	gi11095167	Bacteriophage AR1	gp38	76	34
1282	gi13938232	Homo sapiens	Similar to RIKEN cDNA 2610005H11 gene	78	32
1282	gi13883774	Mycobacterium	NAD-dependent epimerase/dehydratase	83	31

188

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		tuberculosis CDC1551	family protein		
1282	gi5881591	Gallus gallus	homeodomain protein	78	38
1283	gi13938232	Homo sapiens	Similar to RIKEN cDNA 2610005H11 gene	78	32
1283	gi13883774	Mycobacterium tuberculosis CDC1551	NAD-dependent epimerase/dehydratase family protein	83	31
1283	gi5881591	Gallus gallus	homeodomain protein	78	38
1284	gi15779156	Homo sapiens	Similar to RIKEN cDNA 1810073N04 gene	4057	100
1284	gi13097045	Mus musculus	Similar to RIKEN cDNA 1810073N04 gene	1727	91
1284	gi18447388	Drosophila melanogaster	RE05944p	716	32
1285	gi21626874	Drosophila melanogaster	CG9410-PB	354	46
1285	gi7302281	Drosophila melanogaster	CG9410-PA	354	46
1285	gi21166086	Dictyostelium discoideum	Nucleoside diphosphate kinase	164	30
1286	gi20977688	Xenopus laevis	tumorhead	146	33
1286	gi19070822	Mus musculus	Myb protein P42POP	132	29
1286	gi9652255	Ovis aries	DNA binding protein pur-alpha	76	26
1287	gi1006932	Visna virus	envelope polyprotein	61	48
1287	gi6469042	Mus musculus	C184M protein	73	28
1287	gi20988388	Mus musculus	Similar to mammary tumor virus receptor 2	73	28
1288	gi12309630	Homo sapiens	bA438B23.1 (neuronal leucine-rich repeat protein)	319	31
1288	gi6273399	Homo sapiens	melanoma-associated antigen MG50	322	31
1288	gi1504040	Homo sapiens	similar to D.melanogaster peroxidase(U11052)	322	31
1289	gi16769274	Drosophila melanogaster	LD22423p	222	24
1289	gi18700635	Homo sapiens	importin 4	113	23
1289	gi13277562	Homo sapiens	Similar to RIKEN cDNA 8430408O15 gene	113	23
1290	gi21391486	Mus musculus	leucine-rich repeat domain-containing protein	430	43
1290	gi21623740	Rattus norvegicus	Leucine-rich repeat-containing protein 3	425	43
1290	gi21391484	Homo sapiens	leucine-rich repeat domain-containing protein	392	39
1291	gi21624340	Homo sapiens	ceramide kinase	1611	100
1291	gi21624342	Mus musculus	ceramide kinases	1374	86
1291	gi16768660	Drosophila melanogaster	HL01538p	292	41
1292	gi50369	Mus musculus	precursor protein (AA -34 to 244)	204	32
1292	gi312590	Mus musculus	biliary glycoprotein	204	32
1292	gi3549152	Homo sapiens	R29124_1	187	32
1293	gi50369	Mus musculus	precursor protein (AA -34 to 244)	204	32
1293	gi312590	Mus musculus	biliary glycoprotein	204	32
1293	gi3549152	Homo sapiens	R29124_1	187	32
1294	gi21411450	Mus musculus	similar to FLJ00179 protein	1159	91

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1294	gi18676564	Homo sapiens	FLJ00179 protein	993	99
1294	gi17945392	Drosophila melanogaster	RE17452p	486	59
1295	gi7708438	Homo sapiens	dJ885A10.1 (similar to cerebellin precursor)	1020	100
1295	gi5702371	Mus musculus	precerebellin-1	699	70
1295	gi180251	Homo sapiens	precerebellin	696	74
1296	gi3901028	Homo sapiens	neurotensin receptor 2	1436	100
1296	gi1483580	Rattus norvegicus	NTR2 receptor	1073	76
1296	gi17646096	Mus musculus	low affinity neurotensin receptor	1072	77
1298	gi6624583	Homo sapiens	dJ61B2.1 (bullous pemphigoid antigen 1 (230/240kD) isoform 3)	13426	100
1298	gi403124	Homo sapiens	bullous pemphigoid antigen	9121	92
1298	gi15077861	Mus musculus	bullous pemphigoid antigen 1-e	6442	67
1299	gi2114176	Homo sapiens	p97 homologous protein	100	23
1299	gi12654337	Homo sapiens	craniofacial development protein 1	100	23
1299	gi3341899	Homo sapiens	BCNT	100	23
1300	gi6572294	Homo sapiens	bA262A13.1 (novel protein)	499	100
1300	gi21619491	Homo sapiens	similar to expressed sequence AW049604	260	42
1300	gi2460196	Monodelphis domestica	immunoglobulin Igh@ variable domain	65	37
1301	gi18676652	Homo sapiens	FLJ00225 protein	779	100
1301	gi2632952	Bacillus subtilis	yebD	66	51
1301	gi20749947	Drosophila virilis	fruitless class I male isoform	50	40
1302	gi18676652	Homo sapiens	FLJ00225 protein	444	97
1302	gi2632952	Bacillus subtilis	yebD	59	48
1303	gi342299	Macaca fascicularis	preprosomatostatin	226	100
1303	gi338288	Homo sapiens	preprosomatostatin I	226	100
1303	gi21619156	Homo sapiens	somatostatin	226	100
1304	gi14249944	Homo sapiens	Similar to bromodomain-containing 4	109	30
1304	gi2865615	Leishmania peruviana	acidic ribosomal protein P1	93	36
1304	gi343452	Tarsius bancanus	involucrin	114	24
1305	gi219894	Homo sapiens	80K-L protein	124	26
1305	gi187387	Homo sapiens	myristoylated alanine-rich C-kinase substrate	122	26
1305	gi13562004	Nephila madagascariensis	major ampullate spidroin 2-like protein	140	33
1306	gi21744725	Homo sapiens	glycosyl-phosphatidyl-inositol-MAM	1548	48
1306	gi7529597	Homo sapiens	dJ402N21.2 (novel protein with MAM domain)	657	53
1306	gi7529598	Homo sapiens	dJ402N21.3 (novel protein with Immunoglobulin domains)	591	52
1307	gi4455102	Brassica rapa	pollen-specific protein BAN102	72	44
1307	gi4096227	Oryctolagus cuniculus	Ig heavy chain	68	31
1307	gi17017359	Talaromyces emersonii	60S ribosomal protein L2	60	43
1308	gi17429038	Ralstonia	PROBABLE ACYL-COA	1166	56

190

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
		solanacearum	DEHYDROGENASE OXIDOREDUCTASE PROTEIN		
1308	gi9948609	Pseudomonas aeruginosa	probable acyl-CoA dehydrogenase	1121	57
1308	gi13421911	Caulobacter crescentus CB15	acyl-CoA dehydrogenase family protein	1058	54
1309	gi17429038	Ralstonia solanacearum	PROBABLE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE PROTEIN	1166	56
1309	gi9948609	Pseudomonas aeruginosa	probable acyl-CoA dehydrogenase	1121	57
1309	gi13421911	Caulobacter crescentus CB15	acyl-CoA dehydrogenase family protein	1058	54
1310	gi19070124	Mus musculus	zinc transporter-like 3 protein	1087	95
1310	gi20563194	Mus musculus	zinc transporter 6	1075	94
1310	gi9803033	Caenorhabditis elegans	C. elegans TOC-1 protein (corresponding sequence ZC395.3)	279	38
1311	gi854065	Human herpesvirus 6	U88	260	33
1311	gi21928439	Homo sapiens	seven transmembrane helix receptor	174	29
1311	gi18893248	Pyrococcus furiosus DSM 3638	smc-like	177	24
1312	gi5295832	Homo sapiens	dJ21O18.2 (protein similar to collagen)	1142	100
1312	gi6526769	Homo sapiens	HRIHFB2003	1055	97
1312	gi7291408	Drosophila melanogaster	CG11206-PA	738	41
1313	gi19263985	Homo sapiens	Similar to RIKEN cDNA 1300017E09 gene	1565	99
1313	gi19528309	Drosophila melanogaster	LD02310p	573	55
1313	gi7106870	Homo sapiens	HSPC240	227	30
1314	gi22090626	Homo sapiens	HECT domain protein LASU1	1169 0	99
1314	gi6841194	Homo sapiens	HSPC272	9665	99
1314	gi20151907	Drosophila melanogaster	SD03277p	1833	75
1315	gi21542541	Homo sapiens	Similar to HTPAP protein	766	100
1315	gi13182757	Homo sapiens	HTPAP	473	100
1315	gi14020949	Arabidopsis thaliana	phosphatidic acid phosphatase	317	50
1316	gi21542541	Homo sapiens	Similar to HTPAP protein	1204	99
1316	gi13182757	Homo sapiens	HTPAP	915	100
1316	gi14020949	Arabidopsis thaliana	phosphatidic acid phosphatase	460	41
1317	gi180164	Homo sapiens	CD7 antigen protein	1135	93
1317	gi732757	Homo sapiens	CD7 antigen	1135	93
1317	gi14424540	Homo sapiens	CD7 antigen (p41)	1135	93
1319	gi16416764	Homo sapiens	FKSG16	2369	99
1319	gi13905212	Mus musculus	RIKEN cDNA 1200006F02 gene	1833	75
1319	gi14715055	Homo sapiens	Similar to RIKEN cDNA 1110002C08 gene	418	32
1320	gi16416764	Homo sapiens	FKSG16	323	98

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1320	gi13905212	Mus musculus	RIKEN cDNA 1200006F02 gene	257	77
1320	gi14715055	Homo sapiens	Similar to RIKEN cDNA 1110002C08 gene	97	33
1321	gi10834558	Rattus norvegicus	proline arginine-rich end leucine-rich repeat protein	392	32
1321	gi21618473	Homo sapiens	proline arginine-rich end leucine-rich repeat protein	389	32
1321	gi1145773	Homo sapiens	prolargin	389	32
1322	gi20258604	Homo sapiens	sialic acid binding Ig-like lectin 5	1473	84
1322	gi2411475	Homo sapiens	OB binding protein-2	1473	84
1322	gi5759106	Homo sapiens	sialic acid binding Ig-like lectin-5; siglec-5	1473	84
1323	gi20258604	Homo sapiens	sialic acid binding Ig-like lectin 5	1375	87
1323	gi2411475	Homo sapiens	OB binding protein-2	1375	87
1323	gi5759106	Homo sapiens	sialic acid binding Ig-like lectin-5; siglec-5	1375	87
1324	gi20987759	Homo sapiens	Similar to ADAMTS-like 1	886	99
1324	gi15099921	Homo sapiens	ADAM-TS related protein 1	874	98
1324	gi13183078	Homo sapiens	a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3	603	73
1326	gi757915	Homo sapiens	apoCII protein	427	89
1326	gi178836	Homo sapiens	apolipoprotein C-II	427	89
1326	gi342077	Macaca fascicularis	apolipoprotein C-II	371	78
1327	gi21619424	Homo sapiens	Similar to LOC150580	477	100
1327	gi12656449	Plasmodium falciparum	erythrocyte membrane protein 1	63	25
1327	gi15384029	uncultured crenarchaeote 74A4	extracellular protein	64	31
1329	gi16033597	Homo sapiens	SH2 domain-containing phosphatase anchor protein 2d	1003	99
1329	gi16033591	Homo sapiens	SH2 domain-containing phosphatase anchor protein 2b	991	99
1329	gi18092655	Homo sapiens	immunoglobulin superfamily receptor translocation associated protein 3	985	99
1330	gi4877582	Homo sapiens	lipoma HMGIC fusion partner	728	63
1330	gi14272235	Homo sapiens	bA183L8.1 (lipoma HMGIC fusion partner)	445	61
1330	gi15292437	Drosophila melanogaster	LP10272p	187	25
1331	gi17426418	Mus musculus	calmodulin-related protein	788	100
1331	gi12060826	Homo sapiens	serologically defined breast cancer antigen NY-BR-20	610	77
1331	gi5932428	Myxine glutinosa	calmodulin	316	44
1332	gi17862436	Drosophila melanogaster	LD27564p	152	26
1332	gi13311009	Homo sapiens	NYD-SP16	78	26
1333	gi13279251	Homo sapiens	Similar to wingless-related MMTV integration site 6	2000	100
1333	gi11693044	Homo sapiens	WNT6 precursor	2000	100
1333	gi14133265	Homo sapiens	WNT6	2000	100
1334	gi20135611	Homo sapiens	zinc transporter ZnT-5	463	94

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1334	gi19744304	Homo sapiens	zinc transporter 5	463	94
1334	gi19744306	Mus musculus	zinc transporter 5	407	85
1335	gi18480366	Mus musculus	olfactory receptor MOR145-1	310	74
1335	gi21928214	Homo sapiens	seven transmembrane helix receptor	301	77
1335	gi2447219	Homo sapiens	OLF4	295	71
1336	gi20988856	Homo sapiens	protein inhibitor of activated STAT3	3277	100
1336	gi4996563	Homo sapiens	protein inhibitor of activated STAT3	3277	100
1336	gi17149822	Rattus norvegicus	potassium channel regulatory protein KChAP	3211	96
1337	gi4469173	Gallus gallus	delta-9 desaturase	1149	71
1337	gi19908266	Chanos chanos	stearoyl-CoA desaturase	1140	65
1337	gi5738564	Ctenopharyngo don idella	delta-9-desaturase	1132	70
1338	gi14030861	Homo sapiens	paraneoplastic neuronal antigen MA1	1830	99
1338	gi18478557	Rattus norvegicus	paraneoplastic onconeural protein MA1	1752	93
1338	gi15929183	Homo sapiens	modulator of apoptosis 1	990	56
1339	gi5452942	Mus musculus	glucosidase II beta-subunit	134	56
1339	gi163157	Bos taurus	high-mobility-group protein	120	43
1339	gi15076513	Mus musculus	22 kDa neuronal tissue-enriched acidic protein	131	26
1341	gi11177514	Homo sapiens	tandem pore domain potassium channel THIK-2	2234	100
1341	gi11177510	Rattus norvegicus	tandem pore domain potassium channel THIK-2	2215	98
1341	gi15215363	Homo sapiens	potassium channel, subfamily K, member 13	1346	65
1342	gi14336716	Homo sapiens	similar to FBan0003337	1216	100
1342	gi20987336	Mus musculus	RIKEN cDNA A930016P21 gene	427	50
1342	gi19886829	Methanopyrus kandleri AV19	SAM-dependent methyltransferase	104	31
1343	gi19570398	Homo sapiens	hDDM36	1138	43
1343	gi11862939	Mus musculus	DDM36	1134	43
1343	gi11862941	Mus musculus	DDM36E	1125	43
1344	gi21744725	Homo sapiens	glycosyl-phosphatidyl-inositol-MAM	4898	98
1344	gi7529598	Homo sapiens	dJ402N21.3 (novel protein with Immunoglobulin domains)	1548	99
1344	gi7529597	Homo sapiens	dJ402N21.2 (novel protein with MAM domain)	1321	94
1345	gi12276198	Homo sapiens	FKSG40	1020	100
1345	gi12408250	Homo sapiens	FKSG28	1020	100
1345	gi18652934	Xenopus laevis	Mig30	649	49
1346	gi16769552	Drosophila melanogaster	LD38375p	1354	41
1346	gi7523707	Arabidopsis thaliana	Putative membrane protein	1105	39
1346	gi1632829	Plasmodium falciparum	AARP2 protein	467	36
1347	gi20987450	Homo sapiens	LOC146433	1162	95
1347	gi3093373	Mus musculus	small proline-rich protein 2I	64	39
1347	gi912799	Homo sapiens	type I hair keratin	63	33
1348	gi1016012	Rattus norvegicus	neural cell adhesion protein BIG-2 precursor	5093	93
1348	gi19913548	Homo sapiens	similar to axonal-associated cell adhesion	3630	99

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
			molecule		
1348	gi200057	Mus musculus	neuronal glycoprotein	3630	64
1349	gi15292437	Drosophila melanogaster	LP10272p	441	39
1349	gi4877582	Homo sapiens	lipoma HMGIC fusion partner	221	28
1349	gi16648454	Drosophila melanogaster	SD01285p	162	24
1350	gi13097705	Homo sapiens	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	1925	97
1350	gi1340142	Homo sapiens	alpha1-antichymotrypsin	1921	97
1350	gi4165890	Homo sapiens	alpha-1-antichymotrypsin precursor	1850	97
1351	gi21618556	Homo sapiens	trophinin associated protein (tastin)	3134	84
1351	gi905356	Homo sapiens	tastin	3129	84
1351	gi7861746	Mus musculus	GABA-A receptor epsilon-like subunit	165	40
1352	gi12053849	Homo sapiens	DREV protein	1689	100
1352	gi12053851	Homo sapiens	DREV1 protein	1676	99
1352	gi12055091	Mus musculus	DREV protein	1655	97
1353	gi14627081	Homo sapiens	caspase-1 dominant-negative inhibitor Pseudo-ICE	492	100
1353	gi21707335	Homo sapiens	Similar to CARD only protein	462	100
1353	gi186286	Homo sapiens	interleukin 1-beta convertase	445	92
1354	gi17431573	Ralstonia solanacearum	PUTATIVE LIPOPROTEIN TRANSMEMBRANE	82	42
1354	gi995704	Saccharomyces cerevisiae	L3149	69	23
1354	gi1256899	Saccharomyces cerevisiae	Yrl138wp	69	23
1355	gi12034719	Mus musculus	ankyrin-like protein	413	43
1355	gi13469729	Homo sapiens	breast cancer antigen NY-BR-1	415	49
1355	gi21618588	Homo sapiens	testis-specific ankyrin motif containing protein	362	46
1356	gi8272557	Rattus norvegicus	protein kinase WNK1	5439	73
1356	gi6933864	Homo sapiens	kinase deficient protein KDP	3408	100
1356	gi19032238	Homo sapiens	protein kinase WNK3	1664	56
1357	gi8272557	Rattus norvegicus	protein kinase WNK1	5439	73
1357	gi6933864	Homo sapiens	kinase deficient protein KDP	1159	98
1357	gi19032238	Homo sapiens	protein kinase WNK3	530	40
1358	gi10946203	Homo sapiens	neuromedin U receptor 2	785	100
1358	gi9944990	Homo sapiens	neuromedin U receptor-type 2	785	100
1358	gi16877377	Homo sapiens	neuromedin U receptor 2	785	100
1359	gi17861592	Drosophila melanogaster	GH13807p	1234	45
1359	gi18376566	Caenorhabditis elegans	Y105E8A.20	964	49
1359	gi9368514	Leishmania major	methionyl-tRNA synthetase	963	42
1360	gi17389919	Homo sapiens	Similar to major histocompatibility complex, class II, DP beta 1	819	100
1360	gi575494	Homo sapiens	MHC class II lymphocyte antigen beta chain	437	72
1360	gi188479	Homo sapiens	HLA-DPB1	437	72

TABLE 2 A

SEQ ID	Hit ID	Species	Description	S score	Percentage identity
1361	gi3342737	Homo sapiens	R26660_2, partial CDS	1025	97
1361	gi14625940	Homo sapiens	interleukin-10	42	53
1361	gi3005997	okra yellow vein mosaic virus	AC2	77	35
1362	gi3342737	Homo sapiens	R26660_2, partial CDS	1001	94
1362	gi14625940	Homo sapiens	interleukin-10	42	53
1362	gi3005997	okra yellow vein mosaic virus	AC2	77	35
1363	gi13991167	Homo sapiens	sialic acid-binding immunoglobulin-like lectin-like long splice variant	2879	99
1363	gi14625822	Homo sapiens	Siglec-L1	2879	99
1363	gi15824310	Pan troglodytes	sialic acid-binding lectin Siglec-L1	2804	97
1364	gi20072749	Homo sapiens	similar to interferon alpha/beta receptor 1	879	100
1364	gi571296	Homo sapiens	CRFB4	188	27
1364	gi4028135	Gallus gallus	interferon alpha/beta receptor 1	195	27
1365	gi8572055	Homo sapiens	interleukin-1 receptor antagonist homolog 1	823	100
1365	gi6049805	Homo sapiens	interleukin-1 receptor antagonist homolog	823	100
1365	gi6165334	Homo sapiens	interleukin-1-like protein-1	823	100
1366	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2780	40
1366	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	2775	40
1366	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	2774	40
1367	gi4574224	Fundulus heteroclitus	multidrug resistance transporter homolog	287	49
1367	gi19743730	Rattus norvegicus	ATP-binding cassette protein B1b	285	50
1367	gi34525	Homo sapiens	P-glycoprotein (431 AA)	273	50
1368	gi198922	Mus musculus	lymphocyte differentiation antigen	713	100
1368	gi198926	Mus musculus	Ly-6A.2 alloantigen	713	100
1368	gi198930	Mus musculus	differentiation antigen Ly-6E/A	713	100

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
685	gi183150	Homo sapiens	chorionic somatomammotropin CS-5	320	100
685	gi23271170	Homo sapiens	chorionic somatomammotropin hormone 2	275	96
685	gi28188743	Pan troglodytes	placental lactogen PL-B	279	98
686	gi183178	Homo sapiens	hGH-V2	1033	78
686	gi23271170	Homo sapiens	chorionic somatomammotropin hormone 2	707	92
686	gi28188743	Pan troglodytes	placental lactogen PL-B	715	94
688	gi18088830	Homo sapiens	AAH20756	785	100
688	gi183178	Homo sapiens	hGH-V2	1051	79
688	gi30582691	Homo sapiens		785	100
689	gi12653501	Homo sapiens	SERPINF1 protein	2003	95
689	gi30583283	Homo sapiens	, member 1	2003	95
689	gi30585311	synthetic construct	, member 1	2003	95
690	gi20269957	Sus scrofa	AF498759_1 phospholipase C delta 4	1033	88
690	gi21307610	Mus musculus	phospholipase C delta 4	909	77
690	gi571466	Rattus norvegicus	phospholipase C delta-4	893	76
691	gi17864023	Homo sapiens	AF450090_1 KCCR13L	3524	100
691	gi22760385	Homo sapiens	unnamed protein product	3515	99
691	gi22761016	Homo sapiens	unnamed protein product	3524	100
692	gi12697933	Homo sapiens	KIAA1694 protein	3850	100
692	gi20380030	Mus musculus	4933407C03Rik protein	3251	98
692	gi27652547	Homo sapiens	truncated c-Maf-inducing protein	3506	99
693	gi437662	Oryctolagus cuniculus	interleukin-8 receptor subtype B	188	61
693	gi511803	Homo sapiens	interleukin-8 receptor type B	172	57
693	gi576679	Homo sapiens	interleukin 8 receptor B	172	57
694	gi32966069	Homo sapiens	CD39L2 nucleotidase	2514	99
694	gi3335098	Homo sapiens	CD39L2	2520	100
694	gi4691263	Homo sapiens		2513	99
695	gi16566319	Homo sapiens	AF411107_1 G protein-coupled receptor	1843	99
695	gi21928620	Homo sapiens	seven transmembrane helix receptor	1858	100
695	gi22293641	Homo sapiens	putative orphan G protein-coupled receptor 26	845	51
696	gi24660226	Homo sapiens	C-type lectin-like receptor-1	1460	90
696	gi7110216	Homo sapiens	AF200949_1 C-type lectin-like receptor-1	1458	90
696	gi7110218	Mus musculus	AF201457_1 C-type lectin-like receptor 2	322	29
698	gi18089247	Homo sapiens	AAH20966 Similar to ectonucleoside triphosphate diphosphohydrolase 5	2104	100
698	gi30584801	synthetic construct	Homo sapiens ectonucleoside triphosphate diphosphohydrolase 5	2104	100
698	gi3335102	Homo sapiens	CD39L4	2104	100
699	gi804761	Homo sapiens	putative	247	77
700	gi16184225	Drosophila melanogaster	LD24527p	666	42

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
700	gi27447597	Drosophila melanogaster	transcriptional adapter 2S	666	42
700	gi7298997	Drosophila melanogaster	CG9638-PA	666	42
701	gi17225457	Homo sapiens	AF326917_1 autism-related protein 1	1272	36
701	gi27817314	Danio rerio		1234	36
701	gi29468246	Homo sapiens	XTP9	3605	99
702	gi20810589	Homo sapiens	similar to arsenite inducible RNA associated protein	833	99
702	gi22945274	Drosophila melanogaster	CG12795-PA	455	54
702	gi9651711	Mus musculus	AF224494_1 arsenite inducible RNA associated protein	687	80
703	gi13241652	Rattus norvegicus	AF309558_1 supernatant protein factor	2040	93
703	gi13543184	Mus musculus	SEC14-like 2	2038	93
703	gi6624130	Rattus norvegicus	AC004832_1 similar to 45 kDa secretory protein	2150	100
704	gi11066250	Homo sapiens	AF197937_1 presenilins associated rhomboid-like protein	1693	86
704	gi13177766	Homo sapiens	AAH03653 Similar to presenilins associated rhomboid-like protein	1761	99
704	gi15559382	Homo sapiens	AAH14058 presenilins associated rhomboid-like protein	1696	86
705	gi1864091	Rattus norvegicus	PSD-95/SAP90-associated protein-3	4997	95
705	gi2454510	Homo sapiens	PSD-95/SAP90-associated protein-2	2105	47
705	gi6979175	Homo sapiens	AF119818_1 homolog-associated protein 2	2089	47
706	gi11877274	Homo sapiens		2260	99
706	gi21667210	Homo sapiens	AF465765_1 bactericidal/permeability-increasing protein-like 1	2260	99
706	gi21706776	Homo sapiens	Bactericidal/permeability-increasing protein-like 1	2253	99
707	gi16768190	Drosophila melanogaster	GH22974p	647	41
707	gi24659527	Homo sapiens		2006	100
707	gi7291716	Drosophila melanogaster	CG11388-PA	648	41
708	gi14334082	Mus musculus	AF367970_1 thymus LIM protein TLP-A	479	87
708	gi14335908	Mus musculus	thymus LIM protein TLP-A	479	87
708	gi14335909	Mus musculus	thymus LIM protein TLP-B	396	90
709	gi12804105	Homo sapiens	AAH02905 Similar to CG15084 gene product	2090	100
709	gi13649459	Homo sapiens	AF250306_1 putative SB115 protein	2090	100
709	gi18204670	Mus musculus	4930527D15Rik protein	1015	96
710	gi1674440	Homo sapiens	collagen type IV a6 chain	4222	51

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
710	gi1674441	Homo sapiens	collagen type IV a6 chain	4222	51
710	gi556299	Mus musculus	alpha-2 type IV collagen	8126	83
711	gi438007	Gallus gallus	alpha-2-macroglobulin receptor	15742	60
711	gi7861733	Homo sapiens	AF176832_1 low density lipoprotein receptor related protein-deleted in tumor	23654	99
711	gi8926243	Mus musculus	AF270884_1 low density lipoprotein receptor related protein LRP1B/LRP-DIT	23098	92
712	gi17298315	Homo sapiens	candidate tumor suppressor protein	848	100
712	gi7861733	Homo sapiens	AF176832_1 low density lipoprotein receptor related protein-deleted in tumor	848	100
712	gi8926243	Mus musculus	AF270884_1 low density lipoprotein receptor related protein LRP1B/LRP-DIT	731	83
713	gi13544080	Homo sapiens	AAH06171 hypothetical protein MGC2731	1133	100
713	gi20071811	Mus musculus	5830411E10Rik protein	492	55
713	gi33589496	Drosophila melanogaster	LD31278p	401	44
714	gi157409	Drosophila melanogaster	fat protein	3001	40
714	gi22945533	Drosophila melanogaster	CG17941-PA	2292	34
714	gi7295732	Drosophila melanogaster	CG3352-PA	3015	40
715	gi157409	Drosophila melanogaster	fat protein	3007	40
715	gi22945533	Drosophila melanogaster	CG17941-PA	2289	34
715	gi7295732	Drosophila melanogaster	CG3352-PA	3021	40
716	gi17865311	Homo sapiens	AF452102_1 dipeptidyl peptidase-like protein 9	4370	95
716	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	4370	95
716	gi29293087	Homo sapiens	dipeptidyl peptidase 9	4511	95
717	gi2689444	Homo sapiens	ZNF134	1252	57
717	gi31565347	Homo sapiens	LOC284018 protein	1252	57
717	gi9968290	Homo sapiens	zinc finger protein 304	1094	47
718	gi23468368	Mus musculus	1200013F24Rik protein	690	90
718	gi27695305	Mus musculus	1200013F24Rik protein	715	91
718	gi7582294	Homo sapiens	AF208853_1 BM-011	881	100
719	gi1620870	Ciona intestinalis	myoplasmin-C1	410	27
719	gi7416982	Argopecten irradians	myosin heavy chain cardiac muscle specific isoform 1	255	20
719	gi7416983	Argopecten irradians	myosin heavy chain cardiac muscle specific isoform 2	255	20
720	gi13872813	Homo sapiens	fibulin-6	13764	100
720	gi14575679	Homo sapiens	AF156100_1 hemicentin	13720	99
720	gi3879658	Caenorhabditis elegans		1636	29
721	gi13177673	Homo sapiens	AAH03621	1520	45

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
721	gi19354327	Homo sapiens		1520	45
721	gi3822553	Gallus gallus	nuclear calmodulin-binding protein	2238	61
722	gi17223626	Homo sapiens	ATP-binding cassette A10	7963	99
722	gi32350914	Homo sapiens	ATP-binding cassette sub-family A member 10	7943	99
722	gi32350969	Homo sapiens	ATP-binding cassette sub-family A member 10	7943	99
723	gi13374079	Homo sapiens	TAFII140 protein	3677	99
723	gi13374178	Mus musculus	TAFII140 protein	3193	84
723	gi28175603	Homo sapiens	TAF3 protein	2772	99
724	gi17429038	Ralstonia solanacearum	PROBABLE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE PROTEIN	658	61
724	gi22776354	Oceanobacillus iheyensis HTE831	acyl-CoA dehydrogenase	638	63
724	gi28280023	Mus musculus	5730439E10Rik protein	946	85
725	gi21522768	Homo sapiens	unnamed protein product	3060	100
725	gi24047224	Homo sapiens	Similar to EGF-like-domain, multiple 6	3060	100
725	gi6752658	Homo sapiens	AF186084_1 epidermal growth factor repeat containing protein	3055	99
726	gi14530342	Caenorhabditis elegans		1008	36
726	gi6531661	Caenorhabditis elegans	AF195610_1 LIN-41A	1008	36
726	gi6531663	Caenorhabditis elegans	AF195611_1 LIN-41B	1008	36
727	gi1504026	Homo sapiens		5833	99
727	gi22725157	Homo sapiens	minor histocompatibility antigen HA-1	5833	99
727	gi23272016	Homo sapiens	Similar to PTPL1-associated RhoGAP 1	5690	98
728	gi13274120	Homo sapiens		1467	99
728	gi6102996	Mus musculus	Vanin-3	1018	79
728	gi7160973	Homo sapiens	VNN3 protein	1213	96
729	gi27463365	Homo sapiens	a disintegrin-like and metalloprotease with thrombospondin type 1 motifs 9B	8961	99
729	gi28804249	Mus musculus	metalloprotease-disintegrin protease	4974	55
729	gi9581879	Homo sapiens	AF261918_1 disintegrin metalloproteinase with thrombospondin repeats	5723	99
730	gi21063967	Drosophila melanogaster	AT05453p	382	31
730	gi5911409	Drosophila melanogaster	fuzzy	382	31
730	gi7297412	Drosophila melanogaster	CG13396-PA	382	31
731	gi15488017	Homo sapiens	AF407274_1 EW12	2302	100
731	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	2302	100

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
731	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	2302	100
732	gi15488017	Homo sapiens	AF407274_1 EWI2	3200	100
732	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	3200	100
732	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	3200	100
733	gi22266726	Homo sapiens	AF311906_1 LIR-D1 precursor	1303	96
733	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	1303	96
733	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	1303	96
734	gi21748480	Homo sapiens	FLJ00271 protein	605	100
734	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	513	79
734	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	513	79
735	gi31455457	Homo sapiens	putative NFkB activating protein	583	44
735	gi7022838	Homo sapiens	unnamed protein product	1794	99
735	gi7293694	Drosophila melanogaster	CG7323-PA	339	36
736	gi12804169	Homo sapiens	AAH02942	3494	97
736	gi15779178	Homo sapiens	AAH14652 Similar to hypothetical protein BC002942	3532	97
736	gi18088939	Homo sapiens	AAH21143	3494	97
737	gi12836469	Mus musculus	unnamed protein product	3495	87
737	gi26351115	Mus musculus	unnamed protein product	3466	87
737	gi30721603	Mus musculus	RAVER1	3466	87
738	gi12002000	Homo sapiens	AF061732_1 My029 protein	415	100
739	gi15489209	Mus musculus	BC013712 protein	266	31
739	gi21757804	Homo sapiens	unnamed protein product	1226	96
739	gi26354220	Mus musculus	unnamed protein product	1130	79
740	gi15341806	Homo sapiens	AAH13073	2008	100
740	gi19528077	Drosophila melanogaster	AT24025p	165	38
740	gi21627272	Drosophila melanogaster	CG12765-PA	167	24
741	gi23495223	Plasmodium falciparum 3D7	AE014834_50 liver stage antigen, putative	407	23
741	gi32492940	Homo sapiens	medulloblastoma antigen MU-MB-20.201	536	25
741	gi9916	Plasmodium falciparum	liver stage antigen	393	24
742	gi13161060	Homo sapiens	AF332217_1 protocadherin 11	3354	58
742	gi15054521	Homo sapiens	AF217288_1 protocadherin-S	3362	58
742	gi9845485	Homo sapiens	AF169692_1 protocadherin-9	6235	100
743	gi16552038	Homo sapiens	unnamed protein product	2404	99
743	gi21410124	Mus musculus	3230402E02Rik protein	1501	61
743	gi5688958	Homo sapiens	PMMLP	2405	100
744	gi21734445	Rattus norvegicus	BMP/Retinoic acid-inducible neurai-specific protein-2	3987	94
744	gi21734447	Rattus norvegicus	BMP/Retinoic acid-inducible	2948	70

200

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			neural-specific protein-3		
744	gi30348610	Gallus gallus	BMP/retinoic acid-inducible neural-specific protein	2090	52
745	gi2739353	Homo sapiens	ZNF91L	2077	69
745	gi27693081	Homo sapiens		2054	71
745	gi30421228	Homo sapiens	zinc finger protein 430	2486	96
746	gi23272677	Homo sapiens	Similar to zinc finger protein 208	2472	78
746	gi26251755	Homo sapiens	ZNF431 protein	2480	79
746	gi30421228	Homo sapiens	zinc finger protein 430	3174	100
747	gi1212965	Homo sapiens	transmembrane protein	1010	99
747	gi1213221	Rattus norvegicus	transmembrane protein	1006	98
747	gi19683999	Homo sapiens	coated vesicle membrane protein	1010	99
748	gi1199524	Homo sapiens	acid phosphatase	2147	95
748	gi13111975	Homo sapiens	AAH03160 acid phosphatase 2, lysosomal	2143	95
748	gi30584617	synthetic construct	Homo sapiens acid phosphatase 2, lysosomal	2143	95
749	gi15625570	Homo sapiens	AF411981_1 centaurin beta5	3851	95
749	gi28422704	Homo sapiens	CENTB5 protein	2912	100
749	gi30109272	Homo sapiens	CENTB5 protein	4175	99
750	gi10197642	Homo sapiens	AF182421_1 MDS022	647	100
750	gi15929423	Homo sapiens	Hypothetical protein FLJ20502	938	100
750	gi30277696	Mus musculus	D5Buc26e protein	423	78
751	gi18614026	Homo sapiens	zinc finger DNA binding protein p71	998	40
751	gi27693858	Homo sapiens	zinc finger protein 398	998	40
751	gi5630080	Homo sapiens	AC004890_2	984	36
752	gi11345382	Homo sapiens	AF308801_1 vacuolar protein sorting protein 16	3724	95
752	gi12140290	Homo sapiens		3724	95
752	gi15553046	Mus musculus	Vps16	3628	92
753	gi30141048	Homo sapiens	Nogo-66 receptor homolog-1	2226	100
753	gi30141052	Rattus norvegicus	Nogo-66 receptor homolog-1	2130	95
753	gi32351287	Rattus norvegicus	Nogo-66 receptor homolog 2	916	51
754	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2718	39
754	gi25303946	Homo sapiens	alpha-2-macroglobulin	2718	39
754	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	2712	39
755	gi18044501	Mus musculus	angiopoietin-like 3	1692	70
755	gi4929790	Homo sapiens	AF152562_1 angiopoietin-related protein 3	2210	93
755	gi5639997	Mus musculus	AF162224_1 angiopoietin-related protein 3	1692	70
756	gi200057	Mus musculus	neuronal glycoprotein	4821	87
756	gi29837411	Homo sapiens	BIG-2	3898	69
756	gi563133	Rattus norvegicus	BIG-1 protein	4778	87
757	gi16550078	Homo sapiens	unnamed protein product	3710	99
757	gi28175743	Homo sapiens	similar to hypothetical protein FLJ30803	3714	100
757	gi30354720	Mus musculus	AI427653 protein	3609	96
758	gi26329813	Mus musculus	unnamed protein product	3627	93
758	gi28175743	Homo sapiens	similar to hypothetical protein	3612	98

201

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			FLJ30803		
758	gi30354720	Mus musculus	AI427653 protein	3520	95
759	gi21929093	Homo sapiens	seven transmembrane helix receptor	1718	88
759	gi24286029	Homo sapiens	G-protein coupled receptor GPR116	6772	98
759	gi5525078	Rattus norvegicus	seven transmembrane receptor	5048	72
760	gi10440398	Homo sapiens	FLJ00032 protein	1257	61
760	gi11917507	Homo sapiens	HPF1 protein	1254	62
760	gi15929737	Mus musculus	similar to KRAB zinc finger protein KR18	1249	58
761	gi13097633	Homo sapiens	AAH03534 Similar to ATPase, Class I, type 8B, member 1	2325	53
761	gi33440008	Homo sapiens	possible aminophospholipid translocase ATP8B2	3473	66
761	gi3628757	Homo sapiens	FIC1	2576	53
763	gi11558486	Homo sapiens	B-cell lymphoma/leukaemia 11A short form	1314	99
763	gi18089267	Homo sapiens	AAH21098	1153	100
763	gi30410854	Mus musculus		1312	98
764	gi32394378	Homo sapiens	forkhead-associated domain histidine-triad like protein	1808	100
764	gi32394380	Bos taurus	forkhead-associated domain histidine-triad like protein	1638	89
764	gi32394382	Sus scrofa	forkhead-associated domain histidine-triad like protein	1681	91
765	gi31455403	Homo sapiens	aprataxin	241	97
765	gi31455405	Homo sapiens	aprataxin	235	100
765	gi32394378	Homo sapiens	forkhead-associated domain histidine-triad like protein	241	97
766	gi31455403	Homo sapiens	aprataxin	318	100
766	gi32394378	Homo sapiens	forkhead-associated domain histidine-triad like protein	318	100
766	gi32394382	Sus scrofa	forkhead-associated domain histidine-triad like protein	307	93
767	gi26454883	Homo sapiens	hypothetical protein HSPC148	1181	100
767	gi6523797	Homo sapiens	AF110775_1 adrenal gland protein AD-002	1181	100
767	gi6841518	Homo sapiens	AF161497_1 HSPC148	1178	99
768	gi14009597	Homo sapiens	AF282619_1 lysyl oxidase-like 3 protein	1816	98
768	gi14486600	Homo sapiens	AF311313_1 lysyl oxidase-like 3 protein	1816	98
768	gi15186770	Homo sapiens	AF284815_1 lysyl oxidase-like protein	1816	98
769	gi22713410	Homo sapiens	GYLTL1B protein	3229	100
769	gi3954938	Homo sapiens	acetylglucosaminyltransferase-like protein	2292	70
769	gi3954978	Mus musculus	acetylglucosaminyltransferase-like protein	2292	70
770	gi7209721	Mus musculus	DD57	2243	88
770	gi7209723	Homo sapiens	WD-repeat like sequence	2476	99
770	gi8217485	Homo sapiens		2473	99
771	gi16552001	Homo sapiens	unnamed protein product	3169	100

202

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
771	gi18676632	Homo sapiens	FLJ00215 protein	1943	99
771	gi21706685	Mus musculus	9630058J23Rik protein	860	59
772	gi10799166	Homo sapiens	AF305686_1 protein kinase Njmu-R1	1915	99
772	gi32425794	Homo sapiens	NJMU-R1 protein	1888	100
772	gi32450708	Homo sapiens	NJMU-R1 protein	1888	100
773	gi13277972	Mus musculus	phosphatidate cytidyltransferase 2	2286	96
773	gi19344052	Homo sapiens	...	2376	100
773	gi4186023	Homo sapiens	CDS2 protein	2376	100
774	gi17511840	Homo sapiens	AAH18769	2251	99
774	gi20988879	Homo sapiens	Similar to hypothetical gene supported by AL133057; BC018769; BC009436; AL133057; AL133057; AL133057	2251	99
774	gi29387317	Mus musculus	1200011O22Rik protein	1792	79
775	gi13936996	Human herpesvirus 8	ORF73	219	21
775	gi2246532	Human herpesvirus 8	ORF 73, contains large complex repeat CR 73	226	19
775	gi30526291	Saimiriine herpesvirus 2	latency associated nuclear antigen	219	31
776	gi13477379	Homo sapiens	TTYH2 protein	1037	41
776	gi18676664	Homo sapiens	FLJ00231 protein	1796	91
776	gi28422735	Xenopus laevis		1054	40
777	gi16877193	Homo sapiens	AAH16860 G protein-coupled receptor, family C, group 5, member C	939	98
777	gi30583709	Homo sapiens	G protein-coupled receptor, family C, group 5, member C	939	98
777	gi8118032	Homo sapiens	AF207989_1 orphan G-protein coupled receptor	939	98
778	gi15679980	Homo sapiens	C114 protein	930	99
778	gi16769562	Drosophila melanogaster	LD38910p	328	47
778	gi7302978	Drosophila melanogaster	CG8441-PA	328	47
779	gi10726751	Drosophila melanogaster	CG13623-PA	333	53
779	gi21430012	Drosophila melanogaster	GH27470p	333	53
779	gi7406400	Arabidopsis thaliana	putative protein	317	45
780	gi13959018	Homo sapiens	AF361746_1 endothelial cell-selective adhesion molecule	902	100
780	gi13991773	Mus musculus	AF361882_1 endothelial cell-selective adhesion molecule	640	70
780	gi29165726	Mus musculus	Endothelial cell-selective adhesion molecule	640	70
781	gi15422171	Homo sapiens	22 kDa peroxisomal membrane protein 2	1013	100
781	gi297437	Rattus norvegicus	peroxisomal membrane protein	795	76
781	gi8164184	Homo sapiens	22kDa peroxisomal membrane protein-like	1013	100
782	gi7620875	Streptococcus pyogenes	AF232324_1 Sic1.19	203	41

203

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
782	gi7620883	Streptococcus pyogenes	AF232328_1 Sic1.23	203	39
782	gi7621271	Streptococcus pyogenes	AF232522_1 Sic1.217	203	39
783	gi62877	Gallus gallus	type VI collagen alpha-2 subunit preprotein	734	42
783	gi62881	Gallus gallus	type VI collagen subunit alpha2	734	42
783	gi62882	Gallus gallus	type VI collagen subunit alpha2	734	42
784	gi17945608	Drosophila melanogaster	RE26969p	829	48
784	gi7292879	Drosophila melanogaster	CG1998-PA	829	48
784	gi7292910	Drosophila melanogaster	CG11162-PA	597	42
785	gi17066106	Homo sapiens	Novex-3 Titin Isoform	8832	99
785	gi21238650	Calotomus carolinus	titin-like protein	519	62
785	gi27696390	Xenopus laevis	Similar to titin	816	48
786	gi17979434	Arabidopsis thaliana	putative adenylate kinase	193	22
786	gi22136756	Arabidopsis thaliana	putative adenylate kinase	193	22
786	gi30180922	Nitrosomonas europaea ATCC 19718	Adenylate kinase	201	27
787	gi9967224	Macaca fascicularis	hypothetical protein	337	98
788	gi18676610	Homo sapiens	FLJ00204 protein	195	25
788	gi26389725	Mus musculus	unnamed protein product	1390	76
788	gi3002588	Mus musculus	Plenty of SH3s; POSH	197	24
789	gi18676610	Homo sapiens	FLJ00204 protein	250	26
789	gi26329287	Mus musculus	unnamed protein product	1646	75
789	gi26389725	Mus musculus	unnamed protein product	1646	75
790	gi12654107	Homo sapiens	AAH00866	531	88
790	gi13937969	Homo sapiens	TIMP1 protein	531	88
790	gi189382	Homo sapiens	collagenase inhibitor	531	88
791	gi24660226	Homo sapiens	C-type lectin-like receptor-1	1367	90
791	gi7110216	Homo sapiens	AF200949_1 C-type lectin-like receptor-1	1365	90
791	gi7110218	Mus musculus	AF201457_1 C-type lectin-like receptor 2	312	29
792	gi10441350	Mus musculus	olfactory UDP glucuronosyltransferase	1557	68
792	gi4753766	Homo sapiens	UDP glucuronosyltransferase	1593	67
792	gi5802604	Cavia porcellus	UDP glucuronosyltransferase UGT2A3	1781	72
793	gi13325266	Homo sapiens	AAH04450 hypothetical protein MGC2650	888	100
793	gi3688090	Homo sapiens	R32611_2	796	91
793	gi6841228	Homo sapiens	AF161407_1 HSPC289	645	77
794	gi15488645	Mus musculus	methyltransferase Cyt19	1552	76
794	gi18150409	Rattus norvegicus	AF393243_1 methyltransferase	1518	76
794	gi9963861	Homo sapiens	AF226730_1 Cyt19	1729	99
795	gi11877243	Homo sapiens	SSF1/P2Y11 chimeric protein	3802	95
795	gi14602631	Homo sapiens	Peter pan homolog	2080	99
795	gi21619996	Homo sapiens		2080	99

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
796	gi20330550	Homo sapiens	AF251706_1 NK inhibitory receptor precursor	799	98
796	gi30962593	Homo sapiens	AF375481_1 immune receptor expressed on myeloid cells splice variant 2	800	99
796	gi31790204	Homo sapiens	inhibitory receptor IREM1	805	99
797	gi20330550	Homo sapiens	AF251706_1 NK inhibitory receptor precursor	799	98
797	gi30962593	Homo sapiens	AF375481_1 immune receptor expressed on myeloid cells splice variant 2	800	99
797	gi31790204	Homo sapiens	inhibitory receptor IREM1	805	99
798	gi20330550	Homo sapiens	AF251706_1 NK inhibitory receptor precursor	1480	94
798	gi30962591	Homo sapiens	AF375480_1 immune receptor expressed on myeloid cells splice variant 1	1401	93
798	gi31790204	Homo sapiens	inhibitory receptor IREM1	1478	94
799	gi18307481	Homo sapiens	phosphoinositide-binding proteins	2122	100
799	gi27695704	Mus musculus	Connector enhancer of KSR2	678	36
799	gi29691916	Rattus norvegicus	interactor protein for cytohesin exchange factors 1	1651	79
800	gi11493982	Homo sapiens	AF208232_1 TLH29 protein precursor	274	72
800	gi15929988	Homo sapiens	AAH15423 Similar to TLH29 protein precursor	424	89
800	gi21618549	Homo sapiens	TLH29 protein precursor	274	72
801	gi11493982	Homo sapiens	AF208232_1 TLH29 protein precursor	303	70
801	gi15929988	Homo sapiens	AAH15423 Similar to TLH29 protein precursor	445	100
801	gi21618549	Homo sapiens	TLH29 protein precursor	303	70
802	gi12082723	Gallus gallus	AF293805_1 B cell phosphoinositide 3-kinase adaptor	2825	69
802	gi12082725	Mus musculus	AF293806_1 B cell phosphoinositide 3-kinase adaptor	3557	84
802	gi12082811	Gallus gallus	AF315784_1 B cell phosphoinositide 3-kinase adaptor	2330	73
803	gi7959809	Homo sapiens	AF116721_55 PRO1082	545	100
804	gi15384841	Homo sapiens	activating NK receptor	1684	99
804	gi15384843	Homo sapiens	NTB-A receptor	1700	100
804	gi9887089	Mus musculus	AF248635_1 lymphocyte antigen 108 isoform 1	615	43
805	gi10177621	Arabidopsis thaliana	phytoene dehydrogenase-like	195	75
805	gi17979255	Arabidopsis thaliana	AT5g49550/K6M13_10	211	72
805	gi29028742	Arabidopsis thaliana	At5g49550/K6M13_10	211	72
806	gi14270364	Mus musculus	Epigen protein	378	71
806	gi6272269	Rattus norvegicus	NC1 protein	122	52
806	gi7799191	Mus musculus	tomoregulin-1	122	52
807	gi14270364	Mus musculus	Epigen protein	378	71
807	gi6272269	Rattus norvegicus	NC1 protein	122	52

205

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
807	gi7799191	Mus musculus	tomoregulin-1	122	52
808	gi14270364	Mus musculus	Epigen protein	378	71
808	gi6272269	Rattus norvegicus	NC1 protein	122	52
808	gi7799191	Mus musculus	tomoregulin-1	122	52
809	gi27469556	Homo sapiens	Putative neuronal cell adhesion molecule	212	39
809	gi29289929	Danio rerio	neogenin	185	39
809	gi3068592	Mus musculus	punc	198	41
810	gi30348897	Homo sapiens	organic solute transporter beta	643	99
810	gi30348901	Mus musculus	organic solute transporter beta	365	62
811	gi18650584	Homo sapiens	retinoic acid early transcript 1	1070	94
811	gi18650588	Homo sapiens	retinoic acid early transcript 1	1124	99
811	gi21961213	Homo sapiens	UL16 binding protein 2	1070	94
812	gi13872813	Homo sapiens	fibulin-6	485	30
812	gi14575679	Homo sapiens	AF156100_1 hemicentin	485	30
812	gi9280405	Homo sapiens	AF245505_1 adlcan	1372	46
813	gi13872813	Homo sapiens	fibulin-6	861	29
813	gi14575679	Homo sapiens	AF156100_1 hemicentin	857	29
813	gi9280405	Homo sapiens	AF245505_1 adlcan	2436	35
814	gi13872813	Homo sapiens	fibulin-6	861	29
814	gi14575679	Homo sapiens	AF156100_1 hemicentin	857	29
814	gi9280405	Homo sapiens	AF245505_1 adlcan	2436	35
815	gi21619635	Homo sapiens	similar to Alu subfamily SQ sequence contamination warning entry	267	60
815	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	244	62
815	gi6650810	Homo sapiens	AF118094_21 PRO1902	261	63
816	gi12240284	Mus musculus	AF327059_1 apolipoprotein A5	1300	72
816	gi6707433	Homo sapiens	AF202889_1 apolipoprotein A5	1864	100
816	gi6707435	Homo sapiens	AF202890_1 apolipoprotein A5	1864	100
817	gi12240284	Mus musculus	AF327059_1 apolipoprotein A5	1300	72
817	gi6707433	Homo sapiens	AF202889_1 apolipoprotein A5	1864	100
817	gi6707435	Homo sapiens	AF202890_1 apolipoprotein A5	1864	100
818	gi13111784	Homo sapiens	AAH03081 hypothetical protein FLJ10637	1720	99
818	gi13543037	Mus musculus	4933424B01Rik protein	958	80
818	gi14249965	Homo sapiens	AAH08368 hypothetical protein FLJ10637	1724	100
819	gi19344001	Homo sapiens	phospholipase A2, group IID	846	99
819	gi5771420	Homo sapiens	AF112982_1 group IID secretory phospholipase A2	852	100
819	gi6453793	Homo sapiens	AF188625_1 phospholipase A2	846	99
820	gi21751722	Homo sapiens	unnamed protein product	688	84
820	gi26342939	Mus musculus	unnamed protein product	496	59
821	gi11094019	Homo sapiens	AF305057_2 RTS beta	2116	96
821	gi1150421	Homo sapiens	rTSbeta	2122	96

206

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
821	gi12654883	Homo sapiens	AAH01285 rTS beta protein	2122	96
822	gi12803167	Homo sapiens	AAH02387 nucleosome assembly protein 1-like 1	1728	99
822	gi189067	Homo sapiens	NAP	1728	99
822	gi30582885	Homo sapiens	nucleosome assembly protein 1-like 2	1728	99
823	gi13432042	Homo sapiens	integrin-linked kinase-associated serine/threonine phosphatase 2C	2009	99
823	gi16306907	Homo sapiens	AAH06576 integrin-linked kinase-associated serine/threonine phosphatase 2C	2009	99
823	gi20072498	Mus musculus	0710007A14Rik protein	1926	94
824	gi28175169	Mus musculus	1300015B04Rik protein	835	73
824	gi28848867	Homo sapiens	URG11	1164	100
824	gi7768636	Xenopus laevis	Kielin	239	36
825	gi21928259	Homo sapiens	seven transmembrane helix receptor	1023	100
825	gi21928496	Homo sapiens	seven transmembrane helix receptor	1023	100
825	gi21928655	Homo sapiens	seven transmembrane helix receptor	916	89
826	gi18480746	Mus musculus	olfactory receptor MOR261-10	1278	79
826	gi21928655	Homo sapiens	seven transmembrane helix receptor	1456	93
826	gi32052225	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4341246-4340281	1278	79
827	gi4760780	Mus musculus	Ten-m3	364	95
827	gi5307761	Danio rerio	ten-m3	310	78
827	gi6760369	Mus musculus	AF195418_1 ODZ3	364	95
828	gi16265938	Homo sapiens	AF314817_1 FKSG15	2437	98
828	gi21205852	Homo sapiens	AF385429_1 T-cell activation Rho GTPase activating protein; TA-GAP	3756	100
828	gi21205854	Homo sapiens	AF385430_1 T-cell activation Rho GTPase activating protein splice variant 1; TA-GAP	2850	100
829	gi10432396	Homo sapiens		383	62
829	gi30908443	Homo sapiens	CUB and sushi multiple domains 2	388	63
829	gi30908445	Homo sapiens	CUB and sushi multiple domains 3	549	100
830	gi10432396	Homo sapiens		383	62
830	gi30908443	Homo sapiens	CUB and sushi multiple domains 2	388	63
830	gi30908445	Homo sapiens	CUB and sushi multiple domains 3	549	100
831	gi3342148	Chlamydomonas reinhardtii	myosin heavy chain	499	37
831	gi532124	Dictyostelium discoideum	myosin IC	517	41
831	gi8953751	Arabidopsis thaliana	myosin heavy chain MYA2	492	41
832	gi6472600	Chara corallina	unconventional myosin heavy	621	38

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			chain		
832	gi8953751	Arabidopsis thaliana	myosin heavy chain MYA2	621	38
832	gi9453839	Chara corallina	myosin	621	38
834	gi21265163	Homo sapiens		2424	99
834	gi7248845	Homo sapiens	AF231124_1 testican-1	2428	99
834	gi793845	Homo sapiens	testican	2428	99
835	gi20380774	Homo sapiens		2930	99
835	gi22761091	Homo sapiens	unnamed protein product	2350	99
835	gi27502762	Mus musculus	hypothetical protein MGC28931	2712	90
836	gi20380774	Homo sapiens		2946	100
836	gi22761091	Homo sapiens	unnamed protein product	2366	100
836	gi27502762	Mus musculus	hypothetical protein MGC28931	2728	91
837	gi17391348	Homo sapiens	AAH18615 Similar to brain expressed, X-linked 1	664	100
837	gi7689029	Homo sapiens	AF220189_1 uncharacterized hypothalamus protein HBEX2	664	100
837	gi9963771	Homo sapiens	AF183416_1 ovarian granulosa cell 13.0 kDa protein hGR74 homolog	664	100
838	gi15215122	Mus musculus	chondroadherin	428	31
838	gi29571143	Mus musculus	5430427N11Rik protein	430	27
838	gi30908853	Homo sapiens	synleurin	3201	100
839	gi12842465	Mus musculus	unnamed protein product	567	92
839	gi15488920	Homo sapiens	AAH13587 Similar to RIKEN cDNA 2010107G23 gene	632	100
839	gi19354289	Mus musculus	RIKEN cDNA 2010107G23 gene	567	92
840	gi16549697	Homo sapiens	unnamed protein product	2483	99
840	gi20988071	Mus musculus	2600011E07Rik protein	919	80
840	gi21619776	Homo sapiens	Similar to RIKEN cDNA 2600011E07 gene	2491	100
841	gi12963869	Mus musculus	gene trap ankyrin repeat containing protein	223	30
841	gi28565117	Drosophila melanogaster	myosin phosphatase DMBS-S	228	22
841	gi30138665	Nitrosomonas europaea ATCC 19718	Ankyrin-repeat	228	31
842	gi12408272	Homo sapiens	apolipoprotein L-IV splice variant a	1742	100
842	gi12408286	Homo sapiens	apolipoprotein L-IV splice variant a	1742	100
842	gi13374351	Homo sapiens	AF305226_1 apolipoprotein L4	1725	99
843	gi12408272	Homo sapiens	apolipoprotein L-IV splice variant a	1737	99
843	gi12408286	Homo sapiens	apolipoprotein L-IV splice variant a	1737	99
843	gi13374351	Homo sapiens	AF305226_1 apolipoprotein L4	1720	99
844	gi21744725	Homo sapiens	AF478693_1 glycosyl- phosphatidyl-inositol-MAM	2296	100
844	gi25005318	Sus scrofa	MAM domain containing	1804	93

208

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			glycosylphosphatidylinositol anchor 1		
844	gi25005320	Sus scrofa	glycosylphosphatidylinositol anchor 1 protein	1673	92
845	gi21744725	Homo sapiens	AF478693_1 glycosyl-phosphatidyl-inositol-MAM	5051	100
845	gi25005318	Sus scrofa	MAM domain containing glycosylphosphatidylinositol anchor 1	4481	95
845	gi25005320	Sus scrofa	glycosylphosphatidylinositol anchor 1 protein	4350	95
846	gi1066493	Saccharomyces cerevisiae	Ypr144cp	572	30
846	gi32487557	Oryza sativa (japonica cultivar-group)	OSJNBa0013K16.9	565	32
846	gi4007758	Schizosaccharomyces pombe	SPBC1604.06c	613	33
847	gi14280050	Homo sapiens	Vps39/Vam6-like protein	3913	88
847	gi14701768	Homo sapiens	Vam6/Vps39-like protein	3990	89
847	gi23273399	Homo sapiens		4079	98
848	gi23273399	Homo sapiens		4095	99
848	gi25059032	Mus musculus		3128	72
848	gi29467442	Homo sapiens	cytosolic phospholipase A2 delta	1512	41
849	gi14603301	Homo sapiens	Hypothetical protein FLJ11749	986	100
849	gi7291437	Drosophila melanogaster	CG4071-PA	510	49
849	gi9955513	Arabidopsis thaliana	putative protein	340	36
850	gi13161409	Mus musculus	family 4 cytochrome P450	444	73
850	gi13182964	Mus musculus	AF233643_1 cytochrome P450 CYP4F13	196	38
850	gi13278244	Mus musculus	cytochrome P450, family 4, subfamily f, polypeptide 13	196	38
851	gi10944887	Homo sapiens	FGFR-like protein	2475	98
851	gi13183618	Homo sapiens	AF312678_1 FGF homologous factor receptor	2424	97
851	gi13447749	Homo sapiens	AF279689_1 fibroblast growth factor receptor 5	2475	98
852	gi10944887	Homo sapiens	FGFR-like protein	2701	99
852	gi13183618	Homo sapiens	AF312678_1 FGF homologous factor receptor	2650	98
852	gi13447749	Homo sapiens	AF279689_1 fibroblast growth factor receptor 5	2701	99
853	gi10944887	Homo sapiens	FGFR-like protein	583	98
853	gi13183618	Homo sapiens	AF312678_1 FGF homologous factor receptor	583	98
853	gi13447749	Homo sapiens	AF279689_1 fibroblast growth factor receptor 5	583	98
854	gi12667446	Rattus norvegicus	AF336854_1 synaptotagmin VII	2034	95
854	gi6136786	Mus musculus	synaptotagmin VII	2025	95
854	gi643656	Rattus norvegicus	synaptotagmin VII	2034	95
855	gi12053709	Homo sapiens	with thrombospondin type 1 motif, 12	8842	100

209

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
855	gi27817773	Mus musculus	metalloprotease disintegrin 12 protein	7094	80
855	gi5923788	Homo sapiens	AF140675_1 zinc metalloprotease ADAMTS7	2471	51
856	gi15929988	Homo sapiens	AAH15423 Similar to TLH29 protein precursor	179	48
857	gi13542874	Mus musculus	Similar to RIKEN cDNA 2210412D01	1301	74
857	gi17391206	Mus musculus	RIKEN cDNA 2210412D01	1591	94
857	gi28277574	Danio rerio	Similar to RIKEN cDNA 2210412D01 gene	1377	79
858	gi13542874	Mus musculus	Similar to RIKEN cDNA 2210412D01	1301	72
858	gi17391206	Mus musculus	RIKEN cDNA 2210412D01	1591	94
858	gi28277574	Danio rerio	Similar to RIKEN cDNA 2210412D01 gene	1343	79
859	gi20071312	Mus musculus	4933425F03Rik protein	1219	80
859	gi217732	Oryctolagus cuniculus	macrophage scavenger receptor type I subunit	602	38
859	gi33391740	Homo sapiens	MGC45780	1521	98
860	gi20071312	Mus musculus	4933425F03Rik protein	1321	86
860	gi33391740	Homo sapiens	MGC45780	1656	87
860	gi6478784	Mus musculus	scavenger receptor type A SR-A	679	34
861	gi11493463	Homo sapiens	AF130117_38 PRO2852	298	75
861	gi21748687	Homo sapiens	unnamed protein product	319	72
861	gi28801453	Homo sapiens	unnamed protein product	325	77
862	gi14456629	Homo sapiens		1232	50
862	gi15081398	Homo sapiens	AF395541_1 kruppel-like zinc finger protein	1245	54
862	gi29476835	Homo sapiens		1222	47
863	gi16551721	Homo sapiens	unnamed protein product	3124	99
863	gi21320872	Mus musculus	Cog8	2744	87
863	gi7297851	Drosophila melanogaster	CG6488-PA	1143	43
864	gi16307258	Homo sapiens	AAH09717 hypothetical protein	942	100
864	gi22945521	Drosophila melanogaster	CG31922-PA	165	33
864	gi7242597	Homo sapiens	hypothetical protein	942	100
865	gi23274241	Homo sapiens	KIAA1892-like	2039	86
865	gi26332114	Mus musculus	unnamed protein product	1964	82
865	gi26345386	Mus musculus	unnamed protein product	1964	82
866	gi15620885	Homo sapiens	KIAA1913 protein	2495	100
866	gi26339494	Mus musculus	unnamed protein product	2312	90
866	gi28279830	Homo sapiens	KIAA1913 protein	2495	100
867	gi1000448	Rattus norvegicus	Rat kidney AGT2 precursor	2202	81
867	gi12406973	Homo sapiens	alanine-glyoxylate aminotransferase 2	2740	100
867	gi1944136	Rattus norvegicus	beta-alanine-pyruvate aminotransferase	2249	83
868	gi1000448	Rattus norvegicus	Rat kidney AGT2 precursor	1583	84
868	gi12406973	Homo sapiens	alanine-glyoxylate aminotransferase 2	1870	98

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
868	gi1944136	Rattus norvegicus	beta-alanine-pyruvate aminotransferase	1630	86
869	gi26892205	Homo sapiens	1	448	39
869	gi29436673	Mus musculus	1700049K14Rik protein	1732	99
869	gi4165315	Sus scrofa	kallikrein	452	41
870	gi17985046	Brucella melitensis 16M	GLYCOSYL TRANSFERASE	130	28
870	gi20515259	Thermoanaerobacter tengcongensis	predicted glycosyltransferases	133	32
870	gi4455730	Streptomyces coelicolor A3(2)	putative transferase	140	32
872	gi13649477	Homo sapiens	AF250309_1 putative cytokine receptor CRL4 precursor	1998	100
872	gi30584223	synthetic construct	Homo sapiens interleukin 17B receptor	1998	100
872	gi8705222	Homo sapiens	AF212365_1 IL-17B receptor	1998	100
873	gi18676472	Homo sapiens	FLJ00133 protein	6475	100
873	gi20379832	Homo sapiens	FLJ00133 protein	3072	94
873	gi29568116	Mus musculus	secreted protein SST3	3973	84
875	gi14249936	Homo sapiens	AAH08349 Similar to S-adenosylhomocysteine hydrolase-like 1	2581	100
875	gi16588687	Homo sapiens	AF315687_1 S-adenosylhomocysteine hydrolase-like protein	2429	92
875	gi27692283	Mus musculus	S-adenosylhomocysteine hydrolase-like 1	2429	92
876	gi14279990	Homo sapiens	AF294842_1 ubiquitin UBF-fl	458	100
876	gi29791813	Homo sapiens	Ubiquitin-conjugating enzyme E2C, isoform 1	212	74
876	gi30583439	Homo sapiens	ubiquitin-conjugating enzyme E2C	212	74
877	gi20086516	Homo sapiens	AF245303_1 prominin-2 variant A	4241	99
877	gi20086518	Homo sapiens	AF245304_1 prominin-2 variant B	4241	99
877	gi24637566	Rattus norvegicus	prominin-2	3224	74
878	gi29351676	Homo sapiens	Angiopoietin-like 5	2104	100
878	gi29468510	Homo sapiens	putative fibrinogen-like protein	2099	99
878	gi29791750	Homo sapiens	angiopoietin-like 1	392	37
879	gi29351676	Homo sapiens	Angiopoietin-like 5	2100	99
879	gi29468510	Homo sapiens	putative fibrinogen-like protein	2095	99
879	gi29791750	Homo sapiens	angiopoietin-like 1	392	37
880	gi29351676	Homo sapiens	Angiopoietin-like 5	2100	99
880	gi29468510	Homo sapiens	putative fibrinogen-like protein	2095	99
880	gi29791750	Homo sapiens	angiopoietin-like 1	392	37
881	gi11493483	Homo sapiens	AF130117_48 PRO2550	319	66
881	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	303	56
881	gi7770139	Homo sapiens	AF119917_13 PRO1722	318	69
882	gi13543706	Homo sapiens	AAH06003	349	100
882	gi20988061	Mus musculus	1810013D10Rik protein	333	92
882	gi21619079	Homo sapiens		349	100
883	gi11493652	Homo sapiens	AF200708_1 calcium channel	2552	100

211

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			blocker resistance protein CCB1		
883	gi13924720	Homo sapiens	AF252872_1 cystine/glutamate transporter xCT	2552	100
883	gi15082352	Homo sapiens	AAH12087 member 11	2552	100
884	gi14252988	Homo sapiens	SRPK1a protein kinase	2297	86
884	gi23468345	Homo sapiens	SFRS protein kinase 1	2304	87
884	gi507213	Homo sapiens	serine kinase	2297	86
885	gi18044358	Homo sapiens	AAH19883 Similar to lectin-like NK cell receptor	270	57
885	gi9837288	Homo sapiens	C-type lectin	270	57
885	gi9837292	Homo sapiens	C-type lectin	270	57
886	gi22164066	Homo sapiens	AF388385_1 neuroblastoma-amplified protein	7571	99
886	gi30353863	Homo sapiens	NAG protein	7227	99
886	gi4337460	Homo sapiens	neuroblastoma-amplified protein	6886	99
887	gi22164066	Homo sapiens	AF388385_1 neuroblastoma-amplified protein	7309	96
887	gi30353863	Homo sapiens	NAG protein	6965	96
887	gi4337460	Homo sapiens	neuroblastoma-amplified protein	6624	96
888	gi18645094	uncultured proteobacterium	M20/M25/M40 family peptidase, putative	383	38
888	gi19387947	Mus musculus	LOC212933 protein	510	73
888	gi28806353	Vibrio parahaemolyticus	putative M20/M25/M40 family peptidase	387	35
889	gi11558029	Homo sapiens	organic cation transporter	1857	99
889	gi18088251	Homo sapiens	AAH20565 Similar to hBOIT for potent brain type organic ion transporter	1839	95
889	gi9663117	Homo sapiens	organic cation transporter	1849	99
890	gi21732438	Homo sapiens	hypothetical protein	977	100
890	gi26330392	Mus musculus	unnamed protein product	765	80
890	gi26390211	Mus musculus	unnamed protein product	765	80
891	gi13375149	Homo sapiens		853	90
891	gi20072584	Mus musculus	cDNA sequence BC027127	259	37
891	gi7259265	Mus musculus	region	277	47
892	gi16589003	Homo sapiens	AF386649_1 bromodomain-containing 4	6353	99
892	gi18308125	Mus musculus	AF461395_1 bromodomain-containing protein BRD4 long variant	5992	92
892	gi9931486	Mus musculus	AF273217_1 cell proliferation related protein CAP	5994	92
893	gi15420828	Homo sapiens	AF397392_1 NOE3-1	2504	99
893	gi19386926	Rattus norvegicus	AF442822_1 optomedin form B	2484	98
893	gi19386930	Mus musculus	AF442824_1 optomedin form B	2484	98
894	gi22209078	Homo sapiens	hypothetical protein DKFZp566D234	4474	99
894	gi26337809	Mus musculus	unnamed protein product	4135	91
894	gi6330966	Homo sapiens	KIAA1263 protein	4492	100
895	gi12654031	Homo sapiens	AAH00819 Similar to CG6950 gene product	1538	99

212

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
895	gi5002565	Takifugu rubripes	cysteine conjugate beta-lyase	1235	55
895	gi758591	Homo sapiens	glutamine--phenylpyruvate aminotransferase	1193	51
896	gi14017833	Homo sapiens	KIAA1808 protein	2905	99
896	gi21666433	Mus musculus	AF404775_1 actin-binding LIM protein 1 medium isoform	1498	60
896	gi30259308	Mus musculus	actin-binding LIM protein 2	2799	86
897	gi2062399	Rattus norvegicus	protein serine/threonine kinase CPG16	818	52
897	gi6716518	Mus musculus	AF1551 doublecortin-like kinase	818	52
897	gi6716522	Mus musculus	AF155821_1 CPG16	818	52
898	gi2062399	Rattus norvegicus	protein serine/threonine kinase CPG16	818	52
898	gi6716518	Mus musculus	AF1551 doublecortin-like kinase	818	52
898	gi6716522	Mus musculus	AF155821_1 CPG16	818	52
899	gi13436035	Mus musculus	prostaglandin E synthase 2	1583	83
899	gi29179467	Danio rerio	Similar to prostaglandin E synthase 2	1079	60
899	gi9280108	Macaca fascicularis	membrane-associated prostaglandin E synthase-2	1907	97
900	gi12805247	Mus musculus	Complement component 1, q subcomponent, alpha polypeptide	945	70
900	gi20988805	Homo sapiens	complement component 1, q subcomponent, alpha polypeptide	1308	99
900	gi4894854	Homo sapiens	AF135157_1 complement C1q A chain precursor	1308	99
901	gi12841760	Mus musculus	unnamed protein product	928	80
901	gi12846817	Mus musculus	unnamed protein product	931	80
901	gi30802090	Homo sapiens	Similar to RIKEN cDNA 1810059G22 gene	1127	100
902	gi21707458	Homo sapiens	PAX transcription activation domain interacting protein 1 like	2704	87
902	gi2565046	Homo sapiens	CAGF28	3771	97
902	gi4336734	Mus musculus	Pax transcription activation domain interacting protein PTIP	4115	77
903	gi14164561	Xenopus laevis	AF172855_1 Swift	467	79
903	gi4336734	Mus musculus	Pax transcription activation domain interacting protein PTIP	531	93
904	gi15929776	Homo sapiens	AAH15309 growth suppressor 1	135	41
904	gi23271416	Mus musculus	Leprel protein	135	41
904	gi30582917	Homo sapiens	1	135	41
905	gi2443352	Mus musculus	platelet glycoprotein Ib beta	149	45
905	gi30908853	Homo sapiens	synleurin	1549	100
905	gi6808603	Homo sapiens	AF169675_1 leucine-rich repeat transmembrane protein FLRT1	147	40
906	gi13991167	Homo sapiens	sialic acid-binding immunoglobulin-like lectin-like long splice variant	1174	100

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
906	gi14625822	Homo sapiens	AF282256_1 Siglec-L1	1174	100
906	gi23272769	Homo sapiens	SIGLEC-like 1	1174	100
907	gi13435476	Mus musculus	DNA segment, Chr 10, University of California at Los Angeles 1	900	95
907	gi28279553	Danio rerio	Similar to DNA segment, Chr 10, University of California at Los Angeles 1	750	87
907	gi29144983	Mus musculus	DNA segment, Chr 6, ERATO Doi 253, expressed	657	67
908	gi1504040	Homo sapiens		4470	56
908	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	4470	56
908	gi7292259	Drosophila melanogaster	CG12002-PA	2536	36
909	gi1504040	Homo sapiens		4470	56
909	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	4470	56
909	gi7292259	Drosophila melanogaster	CG12002-PA	2536	36
910	gi1504040	Homo sapiens		4112	56
910	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	4112	56
910	gi7292259	Drosophila melanogaster	CG12002-PA	2388	36
911	gi18175295	Homo sapiens	CRB1 isoform II precursor	1258	28
911	gi18182323	Mus musculus	AF406641_1 crumbs-like protein 1 precursor	1242	29
911	gi29144951	Mus musculus	5930402A21 protein	4084	72
912	gi11493463	Homo sapiens	AF130117_38 PRO2852	173	54
912	gi21104464	Homo sapiens	OK/SW-CL.41	184	61
912	gi6650802	Homo sapiens	AF118094_17 PRO1848	200	56
913	gi6808611	Homo sapiens	AF204231_1 88-kDa Golgi protein	3237	99
913	gi6969980	Homo sapiens	AF163441_1 golgin 67	2345	98
913	gi7211438	Homo sapiens	AF164622_1 golgin-67	2327	98
914	gi15030299	Mus musculus	protein kinase, cAMP dependent regulatory, type I beta	1881	94
914	gi200365	Mus musculus	cAMP-dependent protein kinase regulatory subunit	1886	94
914	gi307377	Homo sapiens	cAMP-dependent protein kinase RI-beta regulatory subunit	1957	99
915	gi14017915	Homo sapiens	KIAA1849 protein	3460	100
915	gi7022002	Homo sapiens	unnamed protein product	3074	100
915	gi7022284	Homo sapiens	unnamed protein product	3460	100
916	gi1845577	Mus musculus	-lipoxygenase	2619	77
916	gi30047223	Mus musculus	Arachidonate lipoxygenase, epidermal	2617	77
916	gi3645913	Mus musculus	-lipoxygenase	2619	77
917	gi15489302	Mus musculus	arachidonate lipoxygenase, epidermal	1142	69
917	gi1845577	Mus musculus	-lipoxygenase	1139	69
917	gi30047223	Mus musculus	Arachidonate lipoxygenase,	1142	69

214
TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			epidermal		
918	gi15489302	Mus musculus	arachidonate lipoxygenase, epidermal	1263	75
918	gi1845577	Mus musculus	-lipoxygenase	1260	75
918	gi30047223	Mus musculus	Arachidonate lipoxygenase, epidermal	1263	75
919	gi12053299	Homo sapiens	hypothetical protein	2183	100
919	gi22478033	Homo sapiens	hypothetical protein FLJ22944	3409	91
919	gi22945612	Drosophila melanogaster	CG31652-PA	131	23
920	gi14198207	Mus musculus	hypothetical protein BC008163	1599	98
920	gi19343692	Homo sapiens		1625	100
920	gi7294965	Drosophila melanogaster	CG4452-PA	615	40
921	gi21594983	Homo sapiens	cytokine-like protein C17	238	74
921	gi8132683	Homo sapiens	AF193766_1 cytokine-like protein C17	238	74
922	gi21594983	Homo sapiens	cytokine-like protein C17	238	74
922	gi8132683	Homo sapiens	AF193766_1 cytokine-like protein C17	238	74
923	gi21594983	Homo sapiens	cytokine-like protein C17	381	81
923	gi8132683	Homo sapiens	AF193766_1 cytokine-like protein C17	381	81
924	gi21594983	Homo sapiens	cytokine-like protein C17	263	98
924	gi8132683	Homo sapiens	AF193766_1 cytokine-like protein C17	263	98
925	gi21594983	Homo sapiens	cytokine-like protein C17	591	100
925	gi8132683	Homo sapiens	AF193766_1 cytokine-like protein C17	591	100
926	gi13396317	Homo sapiens		2741	99
926	gi17975777	Homo sapiens	vesicular inhibitory amino acid transporter	2741	99
926	gi31566392	Homo sapiens	Vesicular inhibitory amino acid transporter	2741	99
927	gi22507470	Mus musculus	AI413481 protein	2042	92
927	gi3097285	Rattus norvegicus	ZOG	658	39
927	gi802014	Rattus norvegicus	preadipocyte factor 1	653	39
928	gi16768374	Drosophila melanogaster	GM03282p	357	36
928	gi18088059	Mus musculus	E030025D05Rik protein	1600	89
928	gi6624073	Homo sapiens	AC007743_1 similar to hepatitis delta antigen interacting protein A	1755	93
929	gi14250638	Homo sapiens	AAH08783 Similar to DNA segment, Chr 17, human D6S54E	864	97
929	gi3941733	Mus musculus	AAC82476 BAT4	582	70
929	gi4337106	Homo sapiens	AAD18082 BAT4	864	97
930	gi27476065	Oryza sativa (japonica cultivar-group)	Putative phosphate/phosphoenolpyruvate translocator protein	266	30
930	gi5911433	Rattus norvegicus	AF182714_1 putative phosphate/phosphoenolpyruvate translocator	621	88
930	gi9759107	Arabidopsis thaliana		282	30

215
TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			phosphate/phosphoenolpyruvate translocator protein-like		
931	gi15277895	Homo sapiens	AAH12939 Similar to cardiotrophin-like cytokine; neurotrophin-1/B-cell stimulating factor-3	1204	99
931	gi16356643	Homo sapiens	cardiotrophin-like cytokine	1204	99
931	gi6007643	Homo sapiens	neurotrophin-1/B-cell stimulating factor-3	1204	99
932	gi18490933	Homo sapiens	FLJ21269 protein	846	98
932	gi20268674	Mus musculus	MT-MC1	715	82
932	gi22003732	Homo sapiens	AF527367_1 MTLC	853	99
933	gi15982236	Mus musculus	putative methionyl aminopeptidase	1095	94
933	gi23306398	Arabidopsis thaliana	, putative	744	50
933	gi24899771	Arabidopsis thaliana	, putative	744	50
934	gi1336013	Mus musculus	neurexophilin 2	550	45
934	gi22477181	Homo sapiens	Similar to neurexophilin 4	1649	99
934	gi4104963	Rattus norvegicus	neurexophilin 4	1493	90
935	gi12852913	Mus musculus	unnamed protein product	193	75
935	gi26326067	Mus musculus	unnamed protein product	193	75
937	gi19387136	Homo sapiens	AF479748_1 PYRIN-containing APAF1-like protein 5	874	99
937	gi202806	Rattus norvegicus	vasopressin receptor	561	68
937	gi28436366	Homo sapiens	NALP6	874	99
938	gi11321325	Homo sapiens	AF311862_1 Lin-7b	1030	100
938	gi20381193	Homo sapiens	Lin-7b protein; likely ortholog of mouse LIN-7B; mammalian LIN-7 protein 2	1030	100
938	gi3885828	Rattus norvegicus	lin-7-A	1019	98
939	gi14349125	Homo sapiens	alpha2-glucosyltransferase	738	96
939	gi32490259	Oryza sativa (japonica cultivar-group)	OSJNBb0116K07.1	190	36
939	gi3513451	Rattus norvegicus	potassium channel regulator 1	718	93
940	gi13325140	Homo sapiens	AAH04383	2693	100
940	gi35768	Homo sapiens	polypirimidine tract binding protein	2693	100
940	gi35774	Homo sapiens		2693	100
941	gi21522774	Homo sapiens	unnamed protein product	3068	100
941	gi24047224	Homo sapiens	Similar to EGF-like-domain, multiple 6	3048	99
941	gi6752658	Homo sapiens	AF186084_1 epidermal growth factor repeat containing protein	3043	99
942	gi21522772	Homo sapiens	unnamed protein product	3102	100
942	gi24047224	Homo sapiens	Similar to EGF-like-domain, multiple 6	3043	98
942	gi6752658	Homo sapiens	AF186084_1 epidermal growth factor repeat containing protein	3038	98
943	gi11385648	Homo sapiens	AF273045_1 CTCL tumor antigen sel4-3	3867	99
943	gi17980969	Homo sapiens	AF454056_1 sel4-3r protein	5146	99
943	gi29165763	Mus musculus	3632413B07Rik protein	5213	82

216

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
944	gi13677201	Homo sapiens		2771	100
944	gi17980969	Homo sapiens	AF454056_1 se14-3r protein	3140	99
944	gi29165763	Mus musculus	3632413B07Rik protein	3613	89
945	gi11385648	Homo sapiens	AF273045_1 CTCL tumor antigen se14-3	3806	94
945	gi17980969	Homo sapiens	AF454056_1 se14-3r protein	5085	95
945	gi29165763	Mus musculus	3632413B07Rik protein	5492	85
946	gi11385648	Homo sapiens	AF273045_1 CTCL tumor antigen se14-3	3806	94
946	gi17980969	Homo sapiens	AF454056_1 se14-3r protein	5085	95
946	gi29165763	Mus musculus	3632413B07Rik protein	5566	87
947	gi14043211	Homo sapiens	AAH07594 Similar to RIKEN cDNA 4931428F04 gene	2410	98
947	gi21739633	Homo sapiens	hypothetical protein	2430	97
947	gi25058997	Mus musculus	1110003N12Rik protein	941	63
949	gi19387136	Homo sapiens	AF479748_1 PYRIN-containing APAF1-like protein 5	1735	99
949	gi202806	Rattus norvegicus	vasopressin receptor	1030	64
949	gi28436366	Homo sapiens	NALP6	1735	99
950	gi20338417	Gallus gallus	potassium channel subunit	5079	88
950	gi3875660	Caenorhabditis elegans		2164	45
950	gi3978472	Rattus norvegicus	potassium channel subunit	5376	90
951	gi18147612	Homo sapiens	metalloprotease disintegrin	4376	96
951	gi21908028	Homo sapiens	AF466287_1 a disintegrin and metalloprotease domain 33	4360	96
951	gi21908030	Homo sapiens	a disintegrin and metalloprotease domain 33	4360	96
952	gi12841733	Mus musculus	unnamed protein product	715	92
952	gi18606367	Mus musculus	RIKEN cDNA 4930570C03	715	92
952	gi31581976	Homo sapiens	FLJ20489 protein	472	100
953	gi15420879	Mus musculus	AF398971_1 ankyrin repeat-containing SOCS box protein 10	2049	83
953	gi18031949	Mus musculus	SOCS box protein ASB-18	800	44
953	gi18092200	Homo sapiens	AF417920_1 ASB-10	2174	91
954	gi32707	Homo sapiens	interferon-omega 1	337	51
954	gi386800	Homo sapiens	interferon-alpha	340	51
954	gi491284	synthetic construct	IFN-pseudo-omega 2	799	98
955	gi15928971	Homo sapiens	AAH14951 Similar to neuronal thread protein	430	90
955	gi9844579	Homo sapiens		450	97
955	gi9844580	Homo sapiens		623	84
956	gi11559412	Homo sapiens	NADPH-dependent retinol dehydrogenase/reductase	587	100
956	gi12804321	Homo sapiens	AAH03019 peroxisomal short-chain alcohol dehydrogenase	685	100
956	gi19113668	Homo sapiens	NADP-dependent retinol dehydrogenase short isoform	878	100
957	gi22658418	Mus musculus	cDNA sequence BC030934	1499	68
957	gi28838433	Homo sapiens	DKFZp762A2013 protein	1759	82
957	gi30842594	Homo sapiens	putative sulfhydryl oxidase precursor	1668	78

217

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
958	gi12958660	Homo sapiens	AF321918_1 acid phosphatase	2252	100
958	gi12958663	Homo sapiens	AF321918_4 acid phosphatase variant 3	1285	99
958	gi52871	Mus musculus	lysosomal acid phosphatase	832	45
959	gi11493443	Homo sapiens	AF130117_27 PRO2209	1703	100
959	gi28966	Homo sapiens	alpha 1-antitrypsin	1703	100
959	gi6855601	Homo sapiens	AF113676_1 PRO0684	1703	100
960	gi11493443	Homo sapiens	AF130117_27 PRO2209	2040	95
960	gi177829	Homo sapiens	alpha-1-antitrypsin	2040	95
960	gi28966	Homo sapiens	alpha 1-antitrypsin	2040	95
961	gi11493443	Homo sapiens	AF130117_27 PRO2209	2025	95
961	gi177829	Homo sapiens	alpha-1-antitrypsin	2025	95
961	gi28966	Homo sapiens	alpha 1-antitrypsin	2025	95
962	gi11493443	Homo sapiens	AF130117_27 PRO2209	2036	95
962	gi177829	Homo sapiens	alpha-1-antitrypsin	2036	95
962	gi28966	Homo sapiens	alpha 1-antitrypsin	2036	95
964	gi1841702	Macaca fascicularis	fertilin alpha-I isoform	3138	70
964	gi2632092	Pongo pygmaeus	fertilin alpha protein	4125	92
964	gi794073	Macaca fascicularis	fertilin alpha-I	3138	70
965	gi17887359	Oryctolagus cuniculus	lipophilin AL2	248	54
965	gi4107229	Homo sapiens	lipophilin A	454	100
965	gi4107231	Homo sapiens	lipophilin B	267	60
966	gi13817037	Homo sapiens	E-type ATPase	2812	99
966	gi20988653	Homo sapiens	Similar to ectonucleoside triphosphate diphosphohydrolase 3	2413	99
966	gi3335100	Homo sapiens	CD39L3	2816	100
967	gi180251	Homo sapiens	precerebellin	542	57
967	gi6942096	Mus musculus	CBLN3	936	93
967	gi6942098	Mus musculus	AF218380_1 CBLN3	936	93
968	gi18255724	Mus musculus	LOC215928 protein	131	28
968	gi21750370	Homo sapiens	unnamed protein product	1136	100
968	gi28460663	Rattus norvegicus	Na+ dependent glucose transporter 1	185	30
969	gi21750370	Homo sapiens	unnamed protein product	2545	99
969	gi22328120	Homo sapiens	hypothetical protein DKFZp761N1114	2077	99
969	gi26332881	Mus musculus	unnamed protein product	2116	86
970	gi13161123	Homo sapiens	AF332239_1 transcript Y 10	147	54
970	gi4545317	Acipenser ruthenus	AF129437_1 immunoglobulin light chain precursor	149	25
970	gi9937599	Salmo trutta	AF296378_1 MHC class I heavy chain	153	31
971	gi12964746	Mus musculus	AF316612_1 neuronal pentraxin receptor	2207	88
971	gi2253263	Rattus norvegicus	neuronal pentraxin receptor	2232	89
971	gi4160197	Homo sapiens		2512	99
972	gi27884137	Danio rerio		3553	78
972	gi3170615	Mus musculus	DOC4	4166	96
972	gi4760782	Mus musculus	Ten-m4	4188	96
973	gi14714932	Homo sapiens	AAH10623 -like 1	3770	100
973	gi21748606	Homo sapiens	FLJ00380 protein	3729	96
973	gi541678	Homo sapiens	hbZ17	3729	96

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
974	gi17044301	Leishmania major	possible LIM-binding factor	2875	36
974	gi23095182	Drosophila melanogaster	CG13809-PA	3997	46
974	gi7716100	Rattus norvegicus	AF226993_1 selective LIM binding factor	8413	95
975	gi20799661	Mus musculus	AF503575_1 mucolipin-2	1593	71
975	gi24417793	Mus musculus	mucolipin 2	1593	71
975	gi24417795	Homo sapiens	mucolipin 2	1912	86
976	gi20799661	Mus musculus	AF503575_1 mucolipin-2	2394	83
976	gi24417793	Mus musculus	mucolipin 2	2394	83
976	gi24417795	Homo sapiens	mucolipin 2	2817	99
977	gi1510147	Homo sapiens		309	23
977	gi22477432	Homo sapiens	DKFZP762N2316 protein	4532	91
977	gi403020	Mus musculus	En-2/lacZ fusion protein	988	96
980	gi1513059	Homo sapiens	serin protease with IGF-binding motif	2203	92
980	gi1621244	Homo sapiens	novel serine protease, PRSS11	2203	92
980	gi5281519	Homo sapiens	AF157623_1 HTRA serine protease	2203	92
981	gi11990126	Camelus dromedarius	chymosin	1187	56
981	gi540097	Sus scrofa	preprochymosin	1187	58
981	gi7008025	Callithrix jacchus	prochymosin	1346	64
982	gi27356934	Homo sapiens	extracellular sulfatase SULF-2	293	100
982	gi27356938	Mus musculus	extracellular sulfatase SULF-2	288	100
982	gi28191290	Homo sapiens	sulfatase SULF1 precursor	276	68
984	gi27124671	Homo sapiens	Zn-carboxypeptidase	2008	99
984	gi27529696	Paralichthys olivaceus	carboxypeptidase B	808	49
984	gi6013463	Bothrops jararaca	carboxypeptidase homolog	817	46
985	gi27124671	Homo sapiens	Zn-carboxypeptidase	2008	99
985	gi27529696	Paralichthys olivaceus	carboxypeptidase B	808	49
985	gi6013463	Bothrops jararaca	carboxypeptidase homolog	817	46
986	gi11545705	Homo sapiens	ISCU1	663	99
986	gi11545707	Homo sapiens	ISCU2	845	100
986	gi20381021	Mus musculus	Nifu-pending protein	807	96
987	gi12314022	Homo sapiens		883	89
987	gi22417143	Homo sapiens	CGI-301 protein	853	100
987	gi32879760	Homo sapiens	Snf7 homologue associated with Alix 1	883	89
988	gi12805221	Mus musculus	Lymphocyte antigen 6 complex, locus A	137	33
988	gi198924	Mus musculus	Ly-6A.2	137	33
988	gi201113	Mus musculus	T-cell activation protein	137	33
989	gi17512406	Mus musculus	differential display and activated by p53	1063	67
989	gi25166615	Homo sapiens	AF223000_1 DDA3-like protein	1673	99
989	gi25166621	Homo sapiens	AF322891_1 DDA3-like protein	1673	99
990	gi15990480	Homo sapiens	-binding protein 2	1570	100
990	gi21961217	Homo sapiens	-binding protein 2	1570	100
990	gi22213050	Mus musculus	B230313N05Rik protein	1555	98

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
991	gi204058	Rattus norvegicus	extracellular signal-related kinase 3	1497	62
991	gi23903	Homo sapiens	63kDa protein kinase	2894	99
991	gi27882123	Danio rerio	Similar to mitogen-activated protein kinase 4	1670	61
992	gi17016967	Homo sapiens	AF435011_1 NUANCE	5643	97
992	gi17861384	Homo sapiens	nesprin-2 gamma	5643	97
992	gi24417711	Homo sapiens	nesprin-2	5643	97
993	gi18204756	Mus musculus	2310044D20Rik protein	626	68
993	gi21706580	Mus musculus	A830073O21Rik protein	170	29
993	gi33328302	Homo sapiens	NS5ATP6	997	100
994	gi19353133	Mus musculus	C1q-like	961	66
994	gi26996600	Mus musculus	Similar to C1q-like	1468	94
994	gi32401227	Homo sapiens	AF525315_1 C1q-domain containing protein	1528	98
995	gi14718648	Homo sapiens	allantoicase	1633	99
995	gi20987689	Homo sapiens	Similar to allantoicase	1838	99
995	gi9255889	Mus musculus	AF278712_1 allantoicase	1465	77
996	gi15617341	Homo sapiens	LAG-3 protein precursor	2813	99
996	gi30851187	Homo sapiens	LAG3 protein	1906	99
996	gi579596	Homo sapiens	lymphocyte protein	2651	98
997	gi13810285	Rattus norvegicus	guanine nucleotide release/exchange factor	5813	91
997	gi2522208	Homo sapiens	Ras-GRF2	6407	99
997	gi5882290	Homo sapiens	Ras guanine nucleotide exchange factor 2	6401	99
998	gi22038159	Homo sapiens	AF527605_1 zizimin1	8544	100
998	gi28374168	Mus musculus	AA959601 protein	8001	92
998	gi31419757	Mus musculus	AA959601 protein	8001	92
999	gi10433672	Homo sapiens	unnamed protein product	1530	100
999	gi19263505	Homo sapiens	hypothetical protein FLJ12242	1530	100
999	gi23272394	Homo sapiens	KCTD2 protein	728	67
1000	gi14041697	Homo sapiens		3585	99
1000	gi21594273	Homo sapiens		3626	100
1000	gi25303955	Homo sapiens		3600	100
1001	gi1438532	Rattus norvegicus	rA1	527	25
1001	gi1438534	Rattus norvegicus	rA9	4640	67
1001	gi27371336	Homo sapiens	Similar to CTD-binding SR-like protein rA9	2008	97
1002	gi1438534	Rattus norvegicus	rA9	4640	67
1002	gi27371336	Homo sapiens	Similar to CTD-binding SR-like protein rA9	2008	97
1002	gi7296722	Drosophila melanogaster	CG2926-PA	536	23
1003	gi1675220	Cricetulus griseus	SREBP cleavage activating protein	6194	92
1003	gi23240172	Drosophila melanogaster	CG33131-PA	1077	32
1003	gi30048445	Mus musculus	Similar to SREBP CLEAVAGE-ACTIVATING PROTEIN	2600	89
1004	gi12652851	Homo sapiens	AAH00178 potassium channel modulatory factor	1987	100
1004	gi26453336	Homo sapiens	FIGC1	1983	99

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1004	gi7677058	Homo sapiens	AF155652_1 potassium channel modulatory factor	1983	99
1005	gi26341968	Mus musculus	unnamed protein product	654	54
1005	gi27695389	Mus musculus	MGC58017 protein	1058	98
1005	gi30481648	Homo sapiens		654	54
1006	gi11875318	Mus musculus	synaptotagmin XIII	2004	89
1006	gi14210274	Rattus norvegicus	AF375466_1 synaptotagmin 13	2000	89
1006	gi21410154	Mus musculus	synaptotagmin 13	2004	89
1007	gi11342591	Mus musculus	RanBP7/importin 7	5415	99
1007	gi32330683	Mus musculus	importin 7	5427	99
1007	gi3800881	Homo sapiens	RanBP7/importin 7	5447	100
1008	gi17939650	Homo sapiens	AAH19302 hypothetical protein FLJ12525	3770	99
1008	gi18676522	Homo sapiens	FLJ00158 protein	1512	100
1008	gi27462078	Homo sapiens	AF116730_1 MSTP060	3739	96
1009	gi28981429	Mus musculus	Ddef1 protein	4690	95
1009	gi4063614	Mus musculus	ADP-ribosylation factor-directed GTPase activating protein isoform a	4701	94
1009	gi4406393	Bos taurus	differentiation enhancing factor 1	4700	95
1011	gi13872813	Homo sapiens	fibulin-6	541	29
1011	gi14575679	Homo sapiens	AF156100_1 hemicentin	537	29
1011	gi9280405	Homo sapiens	AF245505_1 adlcan	1631	47
1012	gi12843704	Mus musculus	unnamed protein product	1005	72
1013	gi12833251	Mus musculus	unnamed protein product	710	58
1013	gi17511816	Homo sapiens	AAH18758 Similar to RIKEN cDNA 1110032O22 gene	1468	99
1013	gi20071678	Mus musculus		710	58
1014	gi12833251	Mus musculus	unnamed protein product	748	65
1014	gi17511816	Homo sapiens	AAH18758 Similar to RIKEN cDNA 1110032O22 gene	1288	90
1014	gi20071678	Mus musculus		748	65
1015	gi13529248	Homo sapiens	Centrin 3	839	99
1015	gi2246401	Homo sapiens	centrin	842	100
1015	gi30582215	Homo sapiens		839	99
1016	gi31455256	Homo sapiens	IMAGE3510317 protein	2496	100
1016	gi32492907	Homo sapiens	selenoprotein O	2496	100
1016	gi6572230	Homo sapiens		1879	99
1017	gi31455256	Homo sapiens	IMAGE3510317 protein	2142	100
1017	gi32492907	Homo sapiens	selenoprotein O	2142	100
1017	gi6572230	Homo sapiens		3997	99
1018	gi21928729	Homo sapiens	seven transmembrane helix receptor	2190	99
1018	gi6693701	Homo sapiens	AF147788_1 melanopsin	2226	91
1018	gi6693703	Mus musculus	AF147789_1 melanopsin	1729	74
1019	gi20072741	Mus musculus	E430025L02Rik protein	2634	80
1019	gi28380382	Drosophila melanogaster	CG4168-PA	309	29
1019	gi439296	Homo sapiens	garp	793	37
1020	gi15487302	Homo sapiens	medium-chain acyl-CoA synthetase	1346	99
1020	gi15706421	Homo sapiens	middle-chain acyl-CoA synthetase1	1346	99

221

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1020	gi5019275	Bos taurus	xenobiotic/medium-chain fatty acid:CoA ligase form XL-III	1088	78
1021	gi18874700	Homo sapiens	AF478469_1 Rap1 guanine nucleotide-exchange factor PDZ-GEF2B	5803	98
1021	gi20386206	Homo sapiens	AF478567_1 PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2	5822	98
1021	gi6650766	Homo sapiens	AF117947_1 PDZ domain-containing guanine nucleotide exchange factor I	6216	100
1022	gi18874698	Homo sapiens	AF478468_1 Rap1 guanine nucleotide-exchange factor PDZ-GEF2A	5923	99
1022	gi18874700	Homo sapiens	AF478469_1 Rap1 guanine nucleotide-exchange factor PDZ-GEF2B	5923	99
1022	gi20386206	Homo sapiens	AF478567_1 PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2	5942	100
1023	gi13810306	Homo sapiens	transmembrane protein 7	261	37
1023	gi18250724	Mus musculus	transmembrane protein 7	257	36
1023	gi20270907	Oncorhynchus mykiss	AF483531_1 VHSV-induced protein-5	233	33
1024	gi20071315	Mus musculus	AA589509 protein	1116	76
1024	gi21779866	Mus musculus	AF458068_1 IL-17RE	2052	66
1024	gi21779869	Homo sapiens	AF458069_1 IL-17RE	2896	100
1025	gi20071315	Mus musculus	AA589509 protein	1116	76
1025	gi21779866	Mus musculus	AF458068_1 IL-17RE	2028	72
1025	gi21779869	Homo sapiens	AF458069_1 IL-17RE	2928	100
1026	gi14150450	Rattus norvegicus	AF241241_1 UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T9	1350	93
1026	gi25809274	Homo sapiens	polypeptide N-acetylgalactosaminyltransferase 10	1390	97
1026	gi28268676	Homo sapiens	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10	1384	96
1027	gi15217067	Homo sapiens	AF400436_1 stem cell factor isoform 1	1019	95
1027	gi1827477	Felis catus	stem cell factor	896	84
1027	gi337934	Homo sapiens	stem cell factor	1019	95
1028	gi1377895	Homo sapiens	OB-cadherin-2	1572	56
1028	gi30171995	Homo sapiens	cadherin-24	2721	93
1028	gi30171998	Homo sapiens	cadherin-24 variant	2987	99
1029	gi1377895	Homo sapiens	OB-cadherin-2	1621	60
1029	gi30171995	Homo sapiens	cadherin-24	2770	99
1029	gi30171998	Homo sapiens	cadherin-24 variant	2721	93
1030	gi1398903	Mus musculus	Ca ²⁺ dependent activator protein for secretion	6763	94
1030	gi21541504	Homo sapiens	AF458662_1 calcium-dependent activator protein for	6440	93

222

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			secretion protein		
1030	gi577428	Rattus norvegicus	Ca ²⁺ -dependent activator protein; calcium-dependent actin-binding protein	6449	93
1031	gi11071729	Homo sapiens	putative dipeptidase	1847	99
1031	gi11125344	Homo sapiens	putative metallopeptidase	1319	72
1031	gi32490515	Mus musculus	putative membrane-bound dipeptidase-3	1313	71
1032	gi11493652	Homo sapiens	AF200708_1 calcium channel blocker resistance protein CCB1	2552	100
1032	gi13924720	Homo sapiens	AF252872_1 cystine/glutamate transporter xCT	2552	100
1032	gi15082352	Homo sapiens	AAH12087 member 11	2552	100
1033	gi17028348	Homo sapiens	DKFZP586G1517 protein	3748	100
1033	gi20987924	Mus musculus	2410004L15Rik protein	3473	92
1033	gi29612455	Mus musculus	2410004L15Rik protein	3807	92
1034	gi19352987	Homo sapiens	Similar to KIAA0433 protein	6348	98
1034	gi2887437	Homo sapiens	KIAA0433	6487	99
1034	gi31418648	Mus musculus		4981	97
1035	gi11066463	Rattus norvegicus	AF225961_1 RhoGEF glutamate transport modulator GTRAP48	6385	80
1035	gi19387126	Mus musculus	AF467766_1 guanine nucleotide exchange factor	1778	33
1035	gi7110160	Homo sapiens	guanine nucleotide exchange factor	1792	38
1036	gi10726794	Drosophila melanogaster	CG5521-PA	508	35
1036	gi24061707	Mus musculus	GAP-related interacting partner to E12	986	97
1036	gi4240257	Homo sapiens	KIAA0884 protein	2491	100
1037	gi20269957	Sus scrofa	AF498759_1 phospholipase C delta 4	1472	85
1037	gi21307610	Mus musculus	phospholipase C delta 4	1327	77
1037	gi571466	Rattus norvegicus	phospholipase C delta-4	1295	76
1038	gi16552885	Homo sapiens	unnamed protein product	2084	99
1038	gi26326051	Mus musculus	unnamed protein product	1085	54
1038	gi26327387	Mus musculus	unnamed protein product	1085	54
1039	gi18480186	Mus musculus	olfactory receptor MOR261-6	1323	81
1039	gi32052343	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4384160-4383228	1323	81
1039	gi9368991	Homo sapiens		1410	100
1040	gi29791964	Homo sapiens	Thrombospondin 4	4798	99
1040	gi311626	Homo sapiens	thrombospondin-4	4787	99
1040	gi3860231	Mus musculus	thrombospondin-4	4557	93
1041	gi14043083	Homo sapiens	AAH07524 sperm associated antigen 9	660	100
1041	gi24460121	Homo sapiens	AF327452_1 JNK-associated leucine-zipper protein	273	98
1041	gi29169179	Homo sapiens	PHET	343	98
1042	gi21654741	Homo sapiens	peptide/histidine transporter	2771	95
1042	gi2208839	Rattus norvegicus	peptide/histidine transporter	2344	82

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1042	gi33126130	Homo sapiens	peptide/histidine transporter	2736	94
1043	gi22831474	Drosophila melanogaster	CG14622-PC	2508	47
1043	gi22831475	Drosophila melanogaster	CG14622-PB	2508	47
1043	gi29477075	Mus musculus	Similar to dishevelled associated activator of morphogenesis 1	2521	93
1044	gi15929979	Homo sapiens	AAH15418 Similar to zinc finger protein 345	2476	100
1044	gi33417243	Mus musculus	B230312I18Rik protein	1788	57
1044	gi5080758	Homo sapiens	AC007842_3 BC331191_1	1922	52
1045	gi12655913	Homo sapiens	AF227516_1 sprouty-4A	386	98
1045	gi12655915	Homo sapiens	AF227517_1 sprouty-4C	386	98
1045	gi29747900	Mus musculus	Sprouty homolog 4	320	81
1046	gi29692498	Mus musculus	NAAG-peptidase II	3447	88
1046	gi3211746	Sus scrofa	folylpoly-gamma-glutamate carboxypeptidase	2819	70
1046	gi4539525	Homo sapiens	NAALADase II protein	3881	100
1047	gi21750009	Homo sapiens	unnamed protein product	1414	99
1047	gi23512248	Homo sapiens	Similar to DISCO Interacting Protein 2	676	53
1047	gi26449269	Macaca fascicularis	hypothetical protein	1421	99
1048	gi5918167	Homo sapiens	plexin-B1/SEP receptor	3578	42
1048	gi6651051	Mus musculus	AF133093_2 plexin 6	3147	40
1048	gi9885259	Homo sapiens	AF149019_1 plexin-B3	3140	40
1049	gi15081392	Homo sapiens	AF395817_1 NAC1 protein	1268	55
1049	gi30931339	Mus musculus	Nac1-pending protein	1254	57
1049	gi33392751	Homo sapiens	NAC1 protein	1268	55
1050	gi11692802	Homo sapiens	AF320294_1 ABCG8	3123	99
1050	gi15088540	Homo sapiens	AF324494_1 sterolin-2	3127	99
1050	gi15146444	Homo sapiens	AF351824_1 sterolin-2	3117	99
1051	gi12652851	Homo sapiens	AAH00178 potassium channel modulatory factor	1987	100
1051	gi26453336	Homo sapiens	FIGC1	1983	99
1051	gi7677058	Homo sapiens	AF155652_1 potassium channel modulatory factor	1983	99
1052	gi33395	Homo sapiens		703	70
1052	gi33730	Homo sapiens	immunoglobulin lambda light chain	716	71
1052	gi33734	Homo sapiens	immunoglobulin lambda light chain	716	71
1053	gi21388773	Homo sapiens	kringle-containing protein	1764	80
1053	gi21388775	Homo sapiens	kringle-containing protein	1453	78
1053	gi21623530	Homo sapiens	kringle-containing transmembrane protein	1458	68
1054	gi14495324	Homo sapiens	CMRF35A	432	48
1054	gi18490143	Homo sapiens	CMRF35 leukocyte immunoglobulin-like receptor	432	48
1054	gi396170	Homo sapiens	CMRF-35 antigen	432	48
1055	gi4468255	Homo sapiens	MHC class I antigen	1925	98
1055	gi4468256	Homo sapiens	MHC class I antigen	1974	100
1055	gi487909	Homo sapiens	HLA-A11 antigen A11.1	1914	97
1056	gi21667214	Homo sapiens	AF465767_1	741	100

224

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			bactericidal/permeability-increasing protein-like 3		
1056	gi32490539	Homo sapiens	RY2G5	171	32
1056	gi57732	Rattus rattus	potential ligand-binding protein	210	35
1057	gi21667214	Homo sapiens	AF465767_1 bactericidal/permeability-increasing protein-like 3	2223	99
1057	gi32490539	Homo sapiens	RY2G5	524	31
1057	gi57732	Rattus rattus	potential ligand-binding protein	564	32
1058	gi21667214	Homo sapiens	AF465767_1 bactericidal/permeability-increasing protein-like 3	1916	99
1058	gi32490539	Homo sapiens	RY2G5	434	31
1058	gi57732	Rattus rattus	potential ligand-binding protein	473	33
1059	gi21667214	Homo sapiens	AF465767_1 bactericidal/permeability-increasing protein-like 3	1842	100
1059	gi32490539	Homo sapiens	RY2G5	434	31
1059	gi57732	Rattus rattus	potential ligand-binding protein	473	33
1060	gi13529158	Homo sapiens	AAH05349	1128	99
1060	gi529514	Sus scrofa	neuronal endocrine protein	1092	95
1060	gi7718079	Homo sapiens	neuroendocrine protein 7B2	1148	100
1061	gi15929030	Homo sapiens	AAH14973	2325	100
1061	gi16551493	Homo sapiens	unnamed protein product	2321	99
1061	gi18698601	Homo sapiens	AF467443_1 Smith-Magenis syndrome chromosome region candidate 7 protein	2325	100
1062	gi13543081	Mus musculus	claudin 6	822	70
1062	gi4128041	Homo sapiens	claudin-9 protein	1116	100
1062	gi4325296	Mus musculus	claudin-9	1078	95
1063	gi1215742	Homo sapiens	HIP	434	65
1063	gi14286258	Homo sapiens	AAH08926 ribosomal protein L29	434	65
1063	gi793843	Homo sapiens	ribosomal protein L29	434	65
1064	gi4587895	Rattus norvegicus	AF072509_1 glutamate receptor interacting protein 2	3549	86
1064	gi4731287	Rattus norvegicus	glutamate receptor interacting protein 2	3281	81
1064	gi6601555	Rattus norvegicus	glutamate receptor interacting protein 2	3549	86
1065	gi23496442	Rattus norvegicus	disabled-1	2807	96
1065	gi3288852	Homo sapiens	disabled-1	2865	99
1065	gi8118615	Homo sapiens	AF263547_1 disabled-1	2842	99
1066	gi16877456	Homo sapiens	AAH16974	1711	100
1066	gi20810324	Homo sapiens		1410	86
1066	gi26351033	Mus musculus	unnamed protein product	1236	76
1067	gi15430703	Homo sapiens	AF362953_1 testis specific serine/threonine kinase 2	1858	99
1067	gi2738898	Mus musculus	protein kinase	1683	89
1067	gi33590489	Rattus norvegicus	serine/threonine kinase 22B	1754	92

225

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1068	gi12963879	Homo sapiens	prostaglandin D synthase	980	96
1068	gi13543568	Homo sapiens	PTGDS protein	980	96
1068	gi189772	Homo sapiens	prostaglandin D2 synthase	980	96
1069	gi14336718	Homo sapiens	AE006464_18 similar to HAGH	1157	100
1069	gi20988885	Mus musculus	2810014I23Rik protein	1153	79
1069	gi2459803	Rattus norvegicus	RSP29	645	48
1070	gi13397835	Homo sapiens	annexin A13 isoform b	1795	99
1070	gi21218387	Oryctolagus cuniculus	AF510726_1 annexin XIIIb	1589	88
1070	gi757784	Canis familiaris	annexin XIIIb	1621	89
1071	gi204222	Rattus norvegicus	GABA transporter protein	3094	96
1071	gi21707908	Homo sapiens	, member 1	3126	98
1071	gi31658	Homo sapiens	GABA transporter	3111	98
1072	gi14165176	Rattus norvegicus	AF378093_1 sodium channel beta 3 subunit	823	98
1072	gi7160975	Homo sapiens	voltage-gated sodium channel beta-3 subunit	834	100
1072	gi7161889	Rattus norvegicus	voltage-gated sodium channel beta-3 subunit	823	98
1073	gi20381266	Homo sapiens	Glypican 2	3040	100
1073	gi440127	Rattus norvegicus	cerebroglycan	2506	82
1073	gi5911320	Mus musculus	AF105268_1 glypican-6	1164	44
1074	gi18676470	Homo sapiens	FLJ00132 protein	2515	99
1074	gi19344068	Mus musculus	2700038E08Rik protein	3407	77
1074	gi23274106	Mus musculus	2700038E08Rik protein	3407	77
1075	gi25396387	Homo sapiens	alpha 2,6-sialyltransferase	2844	100
1075	gi27650880	Homo sapiens	beta-galactoside alpha-2,6-sialyltransferase	1183	100
1075	gi452751	Gallus gallus	Gal beta 1,4 GlcNAc alpha 2,6-sialyltransferase	943	54
1076	gi13344995	Homo sapiens	Cat Eye Syndrome critical region protein isoform 1	2002	99
1076	gi13344997	Homo sapiens	Cat Eye Syndrome critical region protein isoform 2	2223	100
1076	gi27503696	Homo sapiens	Similar to cat eye syndrome chromosome region, candidate 5	2223	100
1077	gi13344995	Homo sapiens	Cat Eye Syndrome critical region protein isoform 1	1662	96
1077	gi13344997	Homo sapiens	Cat Eye Syndrome critical region protein isoform 2	1662	96
1077	gi27503696	Homo sapiens	Similar to cat eye syndrome chromosome region, candidate 5	1662	96
1078	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2718	39
1078	gi25303946	Homo sapiens	alpha-2-macroglobulin	2718	39
1078	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	2712	39
1079	gi25303946	Homo sapiens	alpha-2-macroglobulin	1290	35
1079	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	1290	35
1079	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	1291	36
1080	gi25303946	Homo sapiens	alpha-2-macroglobulin	761	31
1080	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	792	32

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1080	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	792	32
1081	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2736	39
1081	gi25303946	Homo sapiens	alpha-2-macroglobulin	2736	39
1081	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	2730	39
1082	gi25303946	Homo sapiens	alpha-2-macroglobulin	1290	35
1082	gi579592	Homo sapiens	alpha 2-macroglobulin 690-730	1290	35
1082	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	1291	36
1083	gi17512361	Mus musculus	esterase 31	2029	66
1083	gi29476863	Mus musculus	Similar to esterase 31	2022	66
1083	gi404389	Mus sp.	carboxylesterase; Es-male	2001	66
1084	gi207286	Rattus norvegicus	TGF-beta masking protein large subunit	8721	89
1084	gi26006334	Mus musculus	latent transforming growth factor beta binding protein 1L	8630	88
1084	gi3493176	Mus musculus	latent TGF beta binding protein	8627	88
1085	gi17985371	Homo sapiens	I3 binding protein	861	100
1085	gi18466808	Homo sapiens	AF283671_1 cervical cancer 1 proto-oncogene-binding protein KG19	853	99
1085	gi21961229	Homo sapiens	BRI3 binding protein	861	100
1086	gi222833	Gallus gallus	M-protein	2924	42
1086	gi2950347	Mus musculus	M-protein	2908	42
1086	gi407097	Homo sapiens	165kD protein	2912	42
1087	gi12655165	Homo sapiens	AAH01438 zinc finger protein 256	693	65
1087	gi30582545	Homo sapiens	zinc finger protein 256	693	65
1087	gi4894364	Homo sapiens	AF067165_1 zinc finger protein 3	693	65
1088	gi1613848	Homo sapiens	zinc finger protein zfp6	311	49
1088	gi30582545	Homo sapiens	zinc finger protein 256	309	56
1088	gi4894364	Homo sapiens	AF067165_1 zinc finger protein 3	309	56
1089	gi12655452	Homo sapiens	keratin associated protein 4.7	981	76
1089	gi12655460	Homo sapiens	keratin associated protein 4.12	970	77
1089	gi12655464	Homo sapiens	keratin associated protein 4.15	973	81
1090	gi12655446	Homo sapiens	keratin associated protein 4.4	400	69
1090	gi12655452	Homo sapiens	keratin associated protein 4.7	383	81
1090	gi12655460	Homo sapiens	keratin associated protein 4.12	400	61
1091	gi12655452	Homo sapiens	keratin associated protein 4.7	1219	90
1091	gi12655460	Homo sapiens	keratin associated protein 4.12	1158	88
1091	gi12655464	Homo sapiens	keratin associated protein 4.15	1260	100
1092	gi15722084	Homo sapiens		1991	100
1092	gi434306	Homo sapiens	lysosomal acid lipase; sterol esterase	1289	63
1092	gi506431	Homo sapiens	lysosomal acid lipase	1289	63
1093	gi15722084	Homo sapiens		1935	100
1093	gi434306	Homo sapiens	lysosomal acid lipase; sterol esterase	1289	63
1093	gi506431	Homo sapiens	lysosomal acid lipase	1289	63
1094	gi20152322	Homo sapiens	putative G-protein coupled receptor	1558	99
1094	gi32526601	Homo sapiens	GPRC5D	1558	99
1094	gi8118040	Homo sapiens	AF209923_1 orphan G-protein	1804	99

227

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			coupled receptor		
1095	gi15099951	Mus musculus	AF384160_1 diacylglycerol acyltransferase 2	596	49
1095	gi18129609	Homo sapiens	AF384161_1 diacylglycerol acyltransferase 2	597	49
1095	gi27693972	Mus musculus	diacylglycerol O-acyltransferase 2	596	49
1096	gi17224598	Homo sapiens	AF293615_1 blood dendritic cell antigen 2 protein	1134	95
1096	gi17225337	Homo sapiens	AF325459_1 dendritic lectin	1134	95
1096	gi17225339	Homo sapiens	AF325460_1 dendritic lectin b isoform	930	80
1097	gi17224598	Homo sapiens	AF293615_1 blood dendritic cell antigen 2 protein	1182	99
1097	gi17225337	Homo sapiens	AF325459_1 dendritic lectin	1182	99
1097	gi17225339	Homo sapiens	AF325460_1 dendritic lectin b isoform	978	84
1098	gi18479834	Mus musculus	olfactory receptor MOR144-1	1220	77
1098	gi21929119	Homo sapiens	seven transmembrane helix receptor	1595	100
1098	gi32063297	Mus musculus	olfactory receptor GA_x6K02T2PVTD-14025733-14026668	1220	77
1099	gi19526645	Homo sapiens	AF430017_1 intestinal membrane mucin MUC17	775	33
1099	gi5911169	Homo sapiens	AF147790_1 transmembrane mucin 12	3049	99
1099	gi5911171	Homo sapiens	AF147791_1 mucin 11	671	54
1100	gi219497	Homo sapiens	biliary glycoprotein	446	34
1100	gi3172151	Homo sapiens	BGPg HUMAN	446	34
1100	gi37198	Homo sapiens	TM1-CEA preprotein	446	34
1101	gi1504040	Homo sapiens		4709	60
1101	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	4709	60
1101	gi7292259	Drosophila melanogaster	CG12002-PA	2660	38
1102	gi1504040	Homo sapiens		4596	59
1102	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	4596	59
1102	gi7292259	Drosophila melanogaster	CG12002-PA	2606	38
1103	gi10435776	Homo sapiens	unnamed protein product	4413	99
1103	gi11611734	Homo sapiens	AF245388_1 GREB1a	510	46
1103	gi7264653	Mus musculus	AF180470_1 Kiaa0575	3121	53
1104	gi16519041	Drosophila melanogaster	AF427496_1 occludin-like protein	184	23
1104	gi20219008	Chlamydomonas reinhardtii	AF394181_1 coiled-coil flagellar protein	673	36
1104	gi7301551	Drosophila melanogaster	CG6059-PA	169	19
1105	gi12654511	Homo sapiens	Torsin family 3, member A	693	96
1105	gi14043167	Homo sapiens	Torsin family 3, member A	693	96
1105	gi15079904	Homo sapiens	Torsin family 3, member A	693	96
1106	gi21666374	Mus musculus	swan	325	72
1106	gi21666376	Mus musculus	swan	325	72

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1106	gi29747798	Mus musculus	3000004N20Rik protein	704	86
1107	gi15076843	Homo sapiens	AF233450_1 pecanex-like protein 1	2759	68
1107	gi18157547	Mus musculus	AF237953_1 pecanex-like 3	4201	93
1107	gi6650377	Mus musculus	AF096286_1 pecanex 1	2767	67
1108	gi15076843	Homo sapiens	AF233450_1 pecanex-like protein 1	2402	73
1108	gi18157547	Mus musculus	AF237953_1 pecanex-like 3	3138	97
1108	gi6650377	Mus musculus	AF096286_1 pecanex 1	2406	73
1109	gi21595759	Homo sapiens	similar to HC6	211	71
1109	gi7020440	Homo sapiens	unnamed protein product	215	57
1109	gi7770237	Homo sapiens	AF119917_62 PRO2822	232	61
1110	gi26333913	Mus musculus	unnamed protein product	749	83
1110	gi26343633	Mus musculus	unnamed protein product	749	83
1110	gi27370621	Homo sapiens	Similar to hypothetical protein FLJ31737	828	95
1111	gi12043567	Homo sapiens	unc-93 related protein	1571	99
1111	gi17390915	Mus musculus	unc93 homolog B	1367	87
1111	gi23271746	Mus musculus	Unc93b protein	1367	87
1112	gi15990461	Homo sapiens	AAH15612 ring finger protein 25	2465	100
1112	gi18490513	Mus musculus	Rnf25 protein	1983	82
1112	gi29179411	Mus musculus	Ring finger protein 25	1988	82
1113	gi19716048	Xenopus laevis	Wee1B kinase	1123	45
1113	gi2827996	Xenopus laevis	wee1 homolog	1291	51
1113	gi644770	Xenopus laevis	Wee1A kinase	1296	51
1115	gi15030119	Mus musculus	3110057O12Rik protein	777	97
1115	gi23093574	Drosophila melanogaster	CG32112-PA	366	42
1115	gi23093575	Drosophila melanogaster	CG32112-PB	397	47
1116	gi11493409	Homo sapiens	AF130117_10 PRO0898	129	59
1116	gi21708029	Homo sapiens	similar to Alu subfamily SQ sequence contamination warning entry	135	70
1116	gi28800991	Homo sapiens	unnamed protein product	124	67
1117	gi13810898	Rattus norvegicus	AF322216_1 inhibin binding protein long isoform	515	32
1117	gi2370143	Homo sapiens	immunoglobulin-like domain-containing 1	503	32
1117	gi2645890	Homo sapiens	IGSF1	503	32
1118	gi2370143	Homo sapiens	immunoglobulin-like domain-containing 1	307	38
1118	gi32330685	Mus musculus	inhibin binding protein/p120 long isoform	310	38
1118	gi32330691	Mus musculus	inhibin binding protein/p120 variant 4	310	38
1119	gi21595190	Mus musculus	2510001A17Rik protein	4878	95
1119	gi21707128	Homo sapiens	Ran binding protein 11	5047	99
1119	gi6650612	Homo sapiens	AF111109_1 Ran binding protein 11	5047	99
1120	gi1399805	Homo sapiens	Bbp/53BP2	2078	46
1120	gi16197705	Homo sapiens	ASPP2 protein	2439	47
1120	gi18652832	Homo sapiens	ASPP1 protein	5703	99

229

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1122	gi2598461	Homo sapiens		1893	97
1122	gi31418316	Homo sapiens	Heat shock 70kD protein binding protein	1893	97
1122	gi4049268	Homo sapiens	putative tumor suppressor ST13	1893	97
1123	gi11991844	Homo sapiens	AF243505_1 fibrocyte-derived protein	676	100
1123	gi12619173	Homo sapiens	melanoma inhibitory activity like protein	676	100
1123	gi12668328	Homo sapiens	melanoma inhibitory activity like protein	676	100
1124	gi22760096	Homo sapiens	unnamed protein product	1047	89
1124	gi27883913	Homo sapiens	POTE	525	46
1124	gi28279813	Homo sapiens	Similar to hypothetical protein DKFZp434A171	743	85
1125	gi11990779	Homo sapiens		548	43
1125	gi22760096	Homo sapiens	unnamed protein product	831	87
1125	gi28279813	Homo sapiens	Similar to hypothetical protein DKFZp434A171	743	85
1126	gi11493483	Homo sapiens	AF130117_48 PRO2550	265	67
1126	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	259	66
1126	gi7770139	Homo sapiens	AF119917_13 PRO1722	266	60
1128	gi16588454	Homo sapiens	AF312374_1 AGTRAP protein	708	95
1128	gi16878260	Homo sapiens	AAH17328 Similar to angiotensin II, type I receptor-associated protein	726	100
1128	gi9621816	Homo sapiens	AF165187_1 ATRAP	708	95
1129	gi12330704	Mus musculus	AF333770_1 cell recognition molecule CASPR4	1376	71
1129	gi17986216	Homo sapiens	AF333769_1 cell recognition molecule CASPR3	1864	98
1129	gi21961652	Mus musculus	contactin associated protein 4	1376	71
1130	gi17986216	Homo sapiens	AF333769_1 cell recognition molecule CASPR3	6812	99
1130	gi18390059	Homo sapiens	AF463518_1 cell recognition protein CASPR4	4738	70
1130	gi21961652	Mus musculus	contactin associated protein 4	4709	68
1131	gi10336504	Homo sapiens	UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase	2014	61
1131	gi21552746	Homo sapiens	AF410457_1 putative polypeptide N-acetylgalactosaminyltransferase	3157	99
1131	gi21552969	Mus musculus	AF467979_1 Williams-Beuren syndrome critical region gene 17	3098	97
1132	gi13625176	Homo sapiens	AF251057_1 thrombospondin	575	46
1132	gi18490857	Homo sapiens	Thrombospondin	575	46
1132	gi31127148	Mus musculus	2610028F08Rik protein	860	96
1133	gi11907599	Homo sapiens	AF208291_1 protein kinase HIPK2	857	50
1133	gi5305331	Mus musculus	AF071070_1 protein kinase Myak-L	856	49
1133	gi5815145	Mus musculus	AF170304_1 nuclear body associated kinase 2b	856	49

230

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1134	gi22267965	Homo sapiens	Similar to KIAA1423 protein	322	100
1134	gi7243227	Homo sapiens	KIAA1423 protein	322	100
1134	gi7300805	Drosophila melanogaster	CG13409-PA	171	51
1135	gi13529338	Mus musculus		1862	48
1135	gi14571502	Homo sapiens	calcium-promoted Ras inactivator	4174	99
1135	gi4185294	Homo sapiens	rasGAP-activating-like protein	1891	48
1137	gi15128103	Mus musculus	AF397007_1 nephronectin	2962	87
1137	gi15128105	Mus musculus	AF397008_1 nephronectin	2934	85
1137	gi15430246	Mus musculus	nephronectin short isoform	2802	83
1138	gi16041675	Homo sapiens	AAH15704 joined to JAZF1	2622	100
1138	gi17862954	Drosophila melanogaster	SD04959p	904	42
1138	gi30046920	Mus musculus	D11Ert530e protein	1941	96
1139	gi12654929	Homo sapiens	AAH01311 mesenchymal stem cell protein DSCD75	719	74
1139	gi17512251	Homo sapiens	AAH19104 mesenchymal stem cell protein DSCD75	716	74
1139	gi7638247	Homo sapiens	AF242773_1 mesenchymal stem cell protein DSCD75	719	74
1140	gi32967231	Homo sapiens	TAFA3	481	100
1140	gi32967237	Homo sapiens	TAFA3.2	923	100
1140	gi32967243	Mus musculus	TAFA3	390	82
1141	gi32967231	Homo sapiens	TAFA3	738	100
1141	gi32967237	Homo sapiens	TAFA3.2	481	100
1141	gi32967243	Mus musculus	TAFA3	634	87
1142	gi10443967	Homo sapiens	AF268610_1 THEG protein	1934	88
1142	gi20306274	Homo sapiens	testicular haploid expressed gene	1934	88
1142	gi7416134	Homo sapiens	testis-specific gene	1934	88
1143	gi21928259	Homo sapiens	seven transmembrane helix receptor	1023	100
1143	gi21928496	Homo sapiens	seven transmembrane helix receptor	1023	100
1143	gi21928655	Homo sapiens	seven transmembrane helix receptor	916	89
1144	gi18480746	Mus musculus	olfactory receptor MOR261-10	1278	79
1144	gi21928655	Homo sapiens	seven transmembrane helix receptor	1456	93
1144	gi32052225	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4341246-4340281	1278	79
1146	gi15779092	Homo sapiens	AAH14613 Similar to syntaxin 18	1295	100
1146	gi30583139	Homo sapiens	syntaxin 18	1295	100
1146	gi30585223	synthetic construct	Homo sapiens syntaxin 18	1295	100
1147	gi14573319	Homo sapiens	AF334755_1 interleukin-1 HY2	812	99
1147	gi14573321	Homo sapiens	AF334756_1 interleukin-1 HY2	812	99
1147	gi18025344	Homo sapiens	interleukin-1 receptor antagonist-like FIL1 theta	809	99
1148	gi1668744	Homo sapiens	HHa5 hair keratin type I	1114	72

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			intermediate filament		
1148	gi3724107	Homo sapiens	type I hair keratin 5	1114	72
1148	gi4103158	Mus musculus	hair keratin acidic 5; Ha5 keratin	1116	72
1149	gi23271416	Mus musculus	Leprel protein	141	30
1149	gi30582917	Homo sapiens	1	139	30
1149	gi6166378	Mus musculus	AF165163_1 growth suppressor 1L	143	30
1150	gi16550754	Homo sapiens	unnamed protein product	1337	90
1150	gi1699265	Homo sapiens	malignant cell expression-enhanced gene/tumor progression-enhanced gene	389	57
1150	gi27529955	Mus musculus	mBB1	1284	86
1151	gi14595019	Homo sapiens	keratin 6 irs	1990	76
1151	gi18031724	Mus musculus	keratin protein K6irs	1948	75
1151	gi27901522	Homo sapiens	keratin 6 irs3	2519	94
1152	gi11066090	Homo sapiens	AF195192_1 matrix metalloprotease MMP-27	2233	84
1152	gi12006364	Tupaia belangeri	AF281673_1 matrix metalloproteinase-27	1859	71
1152	gi3511149	Gallus gallus	matrix metalloproteinase	1213	50
1153	gi11066090	Homo sapiens	AF195192_1 matrix metalloprotease MMP-27	2233	84
1153	gi12006364	Tupaia belangeri	AF281673_1 matrix metalloproteinase-27	1859	71
1153	gi3511149	Gallus gallus	matrix metalloproteinase	1213	50
1154	gi24710913	Homo sapiens	suppressor of fused	2599	100
1154	gi5739507	Homo sapiens	AF175770_1 suppressor of fused	2594	99
1154	gi6689894	Homo sapiens	AF159447_1 Suppressor of Fused	2599	100
1155	gi20387085	Oncorhynchus mykiss	-1	680	31
1155	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability-increasing protein-like 2	2600	100
1155	gi28173296	Cyprinus carpio	bactericidal permeability-increasing protein/lipopolysaccharide-binding protein	702	31
1156	gi12082687	Homo sapiens	Sry-related HMG-box protein	2066	100
1156	gi24047297	Homo sapiens	SRY-box 18	2066	100
1156	gi8894593	Homo sapiens	SOX18 protein	2066	100
1157	gi19526647	Homo sapiens	AF462348_1 oxidored-nitro domain-containing protein	842	92
1157	gi21758574	Homo sapiens	unnamed protein product	922	97
1157	gi7303522	Drosophila melanogaster	CG13178-PA	173	32
1158	gi19526647	Homo sapiens	AF462348_1 oxidored-nitro domain-containing protein	842	92
1158	gi21758574	Homo sapiens	unnamed protein product	922	97
1158	gi7303522	Drosophila melanogaster	CG13178-PA	173	32
1159	gi1794221	Mus musculus	DNA ligase III-beta	2977	89
1159	gi1794223	Mus musculus	DNA ligase III-alpha	2977	89

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1159	gi29165722	Mus musculus	ligase III, DNA, ATP-dependent	3010	89
1160	gi1052871	Homo sapiens	squamous cell carcinoma antigen 2	879	45
1160	gi15667919	Homo sapiens	AF411191_1 SERPINB12	2063	95
1160	gi887465	Homo sapiens	leupin	879	45
1163	gi29611342	Homo sapiens	AF425650_1 MBD1-containing chromatin associated factor	352	52
1163	gi7228149	Mus musculus	ATFa-associated factor	357	29
1163	gi7303705	Drosophila melanogaster	CG12340-PA	187	31
1166	gi14211398	Homo sapiens	AF380342_1 caspase-8L	263	100
1166	gi19401527	Homo sapiens	procaspase-8	223	95
1166	gi20381326	Homo sapiens	Similar to caspase 8, apoptosis-related cysteine protease	263	100
1167	gi10440448	Homo sapiens	FLJ00060 protein	1204	98
1167	gi30466084	Bos taurus	killer cell immunoglobulin-like receptor KIR3DS1	800	53
1167	gi30466086	Bos taurus	killer cell immunoglobulin-like receptor KIR3DL1	783	53
1168	gi1799570	Rattus norvegicus	TIP120	4573	99
1168	gi29792160	Homo sapiens	TIP120 protein	4586	99
1168	gi7688703	Homo sapiens	AF157326_1 TIP120 protein	4573	99
1169	gi13016701	Homo sapiens	activating coreceptor NKp80	1226	100
1169	gi22449867	Macaca fascicularis	NKp80 NK receptor	1122	90
1169	gi7188567	Homo sapiens	AF175206_1 lectin-like receptor F1	1226	100
1171	gi21619190	Homo sapiens	-like 1X-linked	2785	100
1171	gi3021409	Homo sapiens	like 1 protein	3057	100
1171	gi30353941	Homo sapiens	TBL1X protein	3057	100
1172	gi1699265	Homo sapiens	malignant cell expression-enhanced gene/tumor progression-enhanced gene	671	65
1172	gi27529955	Mus musculus	mBB1	646	67
1172	gi33355691	Homo sapiens	transmembrane channel-like protein 4	642	100
1173	gi1699265	Homo sapiens	malignant cell expression-enhanced gene/tumor progression-enhanced gene	671	65
1173	gi27529955	Mus musculus	mBB1	646	67
1173	gi33355691	Homo sapiens	transmembrane channel-like protein 4	642	100
1174	gi16550754	Homo sapiens	unnamed protein product	1881	100
1174	gi1699265	Homo sapiens	malignant cell expression-enhanced gene/tumor progression-enhanced gene	930	81
1174	gi27529955	Mus musculus	mBB1	1810	95
1175	gi13182755	Homo sapiens	AF212237_1 HPHRP	1210	100
1175	gi15929309	Homo sapiens	Phosphotriesterase related	1210	100
1175	gi29791939	Homo sapiens	phosphotriesterase related	1210	100
1177	gi10047271	Homo sapiens	KIAA1598 protein	789	99
1177	gi22539701	Mus musculus	4930506M07Rik protein	818	96
1177	gi26349641	Mus musculus	unnamed protein product	818	96

233

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1178	gi14272704	Homo sapiens	unnamed protein product	157	96
1178	gi19575509	Homo sapiens	unnamed protein product	164	100
1178	gi19575655	Homo sapiens	unnamed protein product	164	100
1182	gi13377880	Cricetulus longicaudatus	AF336043_1 arginine N-methyltransferase p82 isoform	3253	85
1182	gi13377882	Cricetulus longicaudatus	AF336044_1 arginine N-methyltransferase p77 isoform	3253	85
1182	gi13879453	Mus musculus	cDNA sequence BC006705	3260	85
1183	gi14424574	Homo sapiens	AAH09315 phosphatidylserine decarboxylase	777	100
1183	gi16306618	Homo sapiens	AAH01482 phosphatidylserine decarboxylase	1218	96
1183	gi191185	Cricetulus griseus	phosphatidylserine decarboxylase	1128	88
1184	gi10086253	Homo sapiens	glucocorticoid-induced GILZ	460	98
1184	gi11907580	Mus musculus	AF201289_1 TSC22-related inducible leucine zipper 3c	891	87
1184	gi5919161	Homo sapiens	AF183393_1 TSC-22-like Protein	460	98
1185	gi13874437	Homo sapiens	cerebral protein-11	1457	68
1185	gi20987344	Mus musculus	LOC212904 protein	3064	89
1185	gi24980850	Homo sapiens		3283	100
1186	gi14035978	Homo sapiens	unnamed protein product	2577	100
1186	gi14272784	Homo sapiens	unnamed protein product	2577	100
1186	gi16923351	Homo sapiens	AF204270_1 RbBP-35	1431	99
1187	gi18676660	Homo sapiens	FLJ00229 protein	930	97
1187	gi19343701	Mus musculus	RIKEN cDNA A630054L15	913	93
1187	gi25955706	Homo sapiens	Similar to hypothetical protein MGC38041	936	97
1188	gi17865311	Homo sapiens	AF452102_1 dipeptidyl peptidase-like protein 9	4646	100
1188	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	4646	100
1188	gi29293087	Homo sapiens	dipeptidyl peptidase 9	4787	99
1189	gi17865311	Homo sapiens	AF452102_1 dipeptidyl peptidase-like protein 9	4384	95
1189	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	4384	95
1189	gi29293087	Homo sapiens	dipeptidyl peptidase 9	4525	95
1190	gi17865311	Homo sapiens	AF452102_1 dipeptidyl peptidase-like protein 9	4551	98
1190	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	4551	98
1190	gi29293087	Homo sapiens	dipeptidyl peptidase 9	4692	98
1191	gi13097642	Homo sapiens	Ribosomal protein S25	554	99
1191	gi13279149	Homo sapiens	Ribosomal protein S25	554	99
1191	gi13436422	Homo sapiens	Ribosomal protein S25	554	99
1192	gi16549206	Homo sapiens	unnamed protein product	680	100
1193	gi21756739	Homo sapiens	unnamed protein product	4771	97
1193	gi6453538	Homo sapiens	hypothetical protein	4159	99
1193	gi6634025	Homo sapiens	KIAA0379 protein	3467	67
1194	gi12652695	Homo sapiens	AAH00096 HtrA-like serine protease	2116	93
1194	gi5870865	Homo sapiens	serine protease	2116	93

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1194	gi7672669	Homo sapiens	AF141305_1 serine protease Htra2	2116	93
1195	gi1387985	Homo sapiens	A3 adenosine receptor	904	100
1195	gi20988265	Homo sapiens	adenosine A3 receptor	904	100
1195	gi22658481	Homo sapiens	adenosine receptor A3	904	100
1196	gi24078514	Mus musculus	AF454954_1 crossveinless-2	988	91
1196	gi32816043	Mus musculus	BMP-binding endothelial regulator precursor protein	988	91
1196	gi32892146	Homo sapiens	crossveinless-2	1085	100
1197	gi18479346	Mus musculus	olfactory receptor MOR101-1	1334	82
1197	gi18480772	Mus musculus	olfactory receptor MOR101-2	1415	84
1197	gi32054443	Mus musculus	olfactory receptor GA_x6K02T2PB9-2443810-2444775	1415	84
1198	gi16502169	Salmonella enterica subsp. enterica serovar Typhi	putative DNA methylase	751	93
1198	gi29137981	Salmonella enterica subsp. enterica serovar Typhi Ty2	putative DNA methylase	751	93
1198	gi498768	Serratia marcescens	Deoxyadenosyl-methyltransferase	330	51
1199	gi1213589	Xenopus laevis	Prostaglandin D Synthase	290	33
1199	gi16974751	Gallus gallus	CALII	335	37
1199	gi666121	Xenopus laevis	cpl-1	291	33
1200	gi20987993	Mus musculus	MGC41336 protein	1212	90
1200	gi22296200	Thermosynechococcus elongatus BP-1	asparaginyl-tRNA synthetase	1046	46
1200	gi32448516	Pirellula sp.	asparaginyl-tRNA synthetase	1034	47
1201	gi20067381	Homo sapiens	ALMS1 protein	242	41
1201	gi21552774	Mus musculus	AF425257_1 Almstrom syndrome 1 protein	217	38
1201	gi32693320	Homo sapiens	ALMS1 protein	242	41
1202	gi12655061	Homo sapiens	AAH01380	495	92
1202	gi23574788	Macaca fascicularis	succinate dehydrogenase flavoprotein subunit	502	93
1202	gi5759173	Homo sapiens	succinate dehydrogenase flavoprotein subunit	495	92
1203	gi21928186	Mus musculus	GPI-gamma 4; GPIgamma4	1466	61
1203	gi21928188	Mus musculus	GPI-gamma 4; GPIgamma4	1466	61
1203	gi30931171	Mus musculus	GPIgamma4 protein	1466	61
1204	gi15082311	Homo sapiens	AAH12061 -binding protein 3	1534	92
1204	gi9957161	Mus musculus	AF176327_1 alphaCP-3	1708	99
1204	gi9957165	Homo sapiens	AF176329_1 alphaCP-3	1722	100
1205	gi14574118	Caenorhabditis elegans	Dumpy : shorter than wild-type protein 19	233	31
1205	gi16553246	Homo sapiens	unnamed protein product	881	99
1205	gi21739662	Homo sapiens	hypothetical protein	830	95
1206	gi12653341	Homo sapiens	AAH00439 beta	1742	94
1206	gi12804943	Homo sapiens	AAH01924 beta	1742	94
1206	gi31071	Homo sapiens	E-1 beta subunit of the pyruvate dehydrogenase complex	1742	94
1207	gi164851	Oryctolagus	calsequestrin precursor	1908	94

235

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
		cuniculus			
1207	gi2618621	Mus musculus	skeletal muscle calsequestrin	1938	94
1207	gi688292	Homo sapiens	calmitine; calsequestrine	2029	100
1209	gi10432376	Homo sapiens		3334	99
1209	gi11034760	Homo sapiens	NIBAN	3692	99
1209	gi12620192	Homo sapiens	AF288391_1 C1orf24	4775	99
1210	gi2982508	Homo sapiens	TCR beta chain	1290	93
1210	gi3002925	Homo sapiens	T cell receptor beta chain	1277	93
1210	gi3089433	Homo sapiens	T cell receptor beta chain	1028	75
1211	gi12006041	Homo sapiens	AF267857_1 AD038	761	98
1211	gi14189960	Homo sapiens	AF305818_1 PRO0764	141	53
1211	gi33338042	Homo sapiens	AF173896_1 MSTP121	143	46
1213	gi17939498	Homo sapiens	AAH19299 protocadherin gamma subfamily C, 3	4777	99
1213	gi20072790	Homo sapiens	protocadherin gamma subfamily C, 3	4777	99
1213	gi2995719	Homo sapiens	protocadherin 43	4792	100
1214	gi12803363	Homo sapiens	CALR protein	1747	99
1214	gi18088117	Homo sapiens	AAH20493 calreticulin	1747	99
1214	gi30583735	Homo sapiens	calreticulin	1747	99
1215	gi200962	Mus musculus	serine 1 ultra high sulfur protein	254	38
1215	gi200964	Mus musculus	serine 2 ultra high sulfur protein	299	43
1215	gi3228237	Homo sapiens	ultra high sulfur keratin	248	36
1218	gi17223709	Homo sapiens	selenoprotein SelM	235	100
1218	gi17223711	Mus musculus	selenoprotein SelM	188	78
1218	gi26351995	Mus musculus	unnamed protein product	162	76
1221	gi1001963	Homo sapiens	osteopontin	1400	90
1221	gi189151	Homo sapiens	nephropontin precursor	1400	90
1221	gi992950	Homo sapiens	OPN-c	1426	98
1222	gi14326586	Homo sapiens	AF386078_1 serine-cysteine proteinase inhibitor clade C member 1	2252	95
1222	gi179130	Homo sapiens	antithrombin III	2252	95
1222	gi583741	synthetic construct	Antithrombin III	2252	95
1223	gi18088363	Homo sapiens	AAH20669 advanced glycosylation end product-specific receptor	2004	99
1223	gi1841550	Homo sapiens	AAB47491 receptor for advanced glycosylation end products	2004	99
1223	gi561659	Homo sapiens	receptor of advanced glycosylation end products of proteins	2004	99
1224	gi13359193	Homo sapiens	KIAA1660 protein	598	100
1225	gi37231	Homo sapiens	DNA topoisomerase II	8061	99
1225	gi3869382	Homo sapiens	DNA topoisomerase II beta	8048	99
1225	gi790988	Cricetulus longicaudatus		7886	97
1226	gi1881713	Rattus norvegicus	fatty acid transport protein	3039	87
1226	gi20810561	Mus musculus	, member 1	3031	87
1226	gi563829	Mus musculus	fatty acid transport protein	3031	87
1227	gi15080010	Homo sapiens	AAH11789 Similar to COP9	503	44

236

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			complex subunit 7a		
1227	gi15215085	Mus musculus	Cops7b protein	885	71
1227	gi3309176	Mus musculus	COP9 complex subunit 7b	888	71
1228	gi180251	Homo sapiens	precerebellin	544	58
1228	gi6942096	Mus musculus	CBLN3	938	93
1228	gi6942098	Mus musculus	AF218380_1 CBLN3	938	93
1229	gi15620819	Homo sapiens	KIAA1880 protein	2851	99
1229	gi17861952	Drosophila melanogaster	LD01947p	1382	50
1229	gi7291183	Drosophila melanogaster	CG1826-PA	1382	50
1230	gi21756739	Homo sapiens	unnamed protein product	2878	58
1230	gi26354957	Mus musculus	unnamed protein product	5453	95
1230	gi6634025	Homo sapiens	KIAA0379 protein	3166	57
1231	gi20387085	Oncorhynchus mykiss	-1	662	31
1231	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability- increasing protein-like 2	2384	98
1231	gi28173296	Cyprinus carpio	bactericidal permeability- increasing protein/lipopolysaccharide- binding protein	680	31
1232	gi20387085	Oncorhynchus mykiss	-1	654	31
1232	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability- increasing protein-like 2	2389	99
1232	gi28173296	Cyprinus carpio	bactericidal permeability- increasing protein/lipopolysaccharide- binding protein	672	30
1233	gi20387085	Oncorhynchus mykiss	-1	688	31
1233	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability- increasing protein-like 2	2595	99
1233	gi28173296	Cyprinus carpio	bactericidal permeability- increasing protein/lipopolysaccharide- binding protein	710	31
1234	gi18257341	Mus musculus	Expressed sequence AW060207	2106	69
1234	gi2191168	Arabidopsis thaliana	contains similarity to myosin heavy chain	207	26
1234	gi2879804	Schizosaccharomyces pombe	SPAC23A1.16c	163	28
1235	gi11493528	Homo sapiens	AF130117_71 PRO1953	671	100
1236	gi21754036	Homo sapiens	unnamed protein product	998	99
1236	gi30411057	Mus musculus	RIKEN cDNA B230219D22	954	93
1236	gi31565787	Homo sapiens	FLJ37562 protein	1002	100
1237	gi27469556	Homo sapiens	Putative neuronal cell adhesion molecule	3516	99
1237	gi3068592	Mus musculus	punc	2976	86
1237	gi4206390	Homo sapiens	putative neuronal cell adhesion	1569	98

237

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			molecule		
1238	gi12667401	Homo sapiens	AF326731_1 NUF2R	2347	99
1238	gi14317902	Homo sapiens	kinetochore protein Nuf2	2347	99
1238	gi18043223	Mus musculus	NUF2R protein	1754	73
1239	gi10435493	Homo sapiens	unnamed protein product	2702	99
1239	gi7022901	Homo sapiens	unnamed protein product	3682	99
1239	gi7688176	Homo sapiens	hypothetical protein	3688	99
1240	gi21634823	Homo sapiens	AF389428_1 semaphorin 6D isoform 3	5142	91
1240	gi21634825	Homo sapiens	AF389429_1 semaphorin 6D isoform 4	5667	98
1240	gi21634827	Homo sapiens	AF389430_1 semaphorin 6D isoform 1	3112	63
1241	gi14036200	Homo sapiens	unnamed protein product	245	97
1243	gi21671105	Homo sapiens	RAD52B	1134	100
1243	gi23468352	Homo sapiens	Similar to RAD52B	963	99
1243	gi32967621	Mus musculus	2410008M22Rik protein	828	74
1244	gi15928404	Mus musculus	Fasting-inducible integral membrane protein TM6P1	185	36
1244	gi18490578	Mus musculus	A630041N19 protein	449	71
1244	gi20379926	Mus musculus	Fasting-inducible integral membrane protein TM6P1	185	36
1245	gi18490578	Mus musculus	A630041N19 protein	875	70
1245	gi29792229	Homo sapiens	FLJ90024 protein	297	33
1245	gi6013381	Rattus norvegicus	AF186469_1 TM6P1	296	33
1246	gi28626251	Homo sapiens	calcium-permeable store-operated channel TRPM3c	1194	100
1246	gi28626253	Homo sapiens	calcium-permeable store-operated channel TRPM3d	1194	100
1246	gi28626255	Homo sapiens	calcium-permeable store-operated channel TRPM3e	1194	100
1247	gi17386053	Mus musculus	AF444274_1 Jedi protein	2269	50
1247	gi18044366	Homo sapiens	AAH20198 Similar to MEGF10 protein	3468	99
1247	gi18252658	Mus musculus	AF461685_1 Jedi-736 protein	2269	50
1248	gi20987880	Mus musculus	E130103I17Rik protein	3580	87
1248	gi28204917	Mus musculus	E130103I17Rik protein	3801	86
1248	gi4588087	Homo sapiens	AF095771_1 PTH-responsive osteosarcoma B1 protein	4080	94
1249	gi13591434	Homo sapiens		1160	100
1249	gi13591435	Homo sapiens		976	99
1249	gi19913471	Homo sapiens		1265	99
1250	gi16605581	Homo sapiens	H-rev107-like protein 5	1451	100
1250	gi21707989	Homo sapiens	Similar to H-rev107-like protein 5	1382	96
1250	gi6048565	Homo sapiens	AF092922_1 retinoid inducible gene 1	376	54
1251	gi21263094	Rattus norvegicus	AF512430_1 tramdorin 1	1665	81
1251	gi27924388	Mus musculus	Tramdorin 1	1668	82
1251	gi31871293	Homo sapiens	proton/amino acid transporter 2	2010	99
1252	gi14571904	Rattus norvegicus	AF361239_1 lysosomal amino acid transporter 1	1931	78
1252	gi31324239	Homo sapiens	proton-coupled amino acid transporter	2174	90

238

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1252	gi31871291	Homo sapiens	proton/amino acid transporter 1	2195	90
1254	gi1563885	Homo sapiens	fibroblast growth factor homologous factor 1	917	90
1254	gi1669500	Mus musculus	fibroblast growth factor homologous factor 1	917	90
1254	gi20988932	Mus musculus	Fgf12 protein	916	98
1255	gi19263005	Ciona intestinalis	leucine-rich repeat dynein light chain	759	75
1255	gi2760161	Anthocidaris crassispina	outer arm dynein light chain 2	658	67
1255	gi7303901	Drosophila melanogaster	CG8800-PA	554	58
1256	gi12666529	Mus musculus	b,b-carotene-9',10'-dioxygenase	2356	80
1256	gi12666531	Homo sapiens	putative b,b-carotene-9',10'-dioxygenase	2982	99
1256	gi14582265	Homo sapiens	AF276432_1 putative carotene dioxygenase	2918	99
1257	gi12666529	Mus musculus	b,b-carotene-9',10'-dioxygenase	2305	81
1257	gi12666531	Homo sapiens	putative b,b-carotene-9',10'-dioxygenase	2850	96
1257	gi14582265	Homo sapiens	AF276432_1 putative carotene dioxygenase	2786	95
1258	gi15559697	Homo sapiens	AAH14205 Similar to neural cell adhesion molecule 1	157	28
1258	gi28703938	Homo sapiens	Similar to neural cell adhesion molecule 1	157	28
1258	gi61	Bos taurus	calmodulin-independent adenylate cyclase	158	28
1260	gi1079734	Mus musculus	citron	1291	94
1260	gi2745840	Rattus norvegicus	postsynaptic density protein; citron	1262	93
1260	gi3599509	Mus musculus	rho/rac-interacting citron kinase	1286	94
1261	gi28277755	Danio rerio	proteinase inhibitor, clade E, member 2	479	30
1261	gi28435507	Sus scrofa	nexin-1	467	30
1261	gi32485107	Homo sapiens	nexin-related serine protease inhibitor	2002	92
1262	gi13383364	Homo sapiens	claudin-1	223	97
1262	gi15214678	Homo sapiens	AAH12471 claudin 1	223	97
1262	gi7381083	Homo sapiens	AF134160_1 claudin-1	223	97
1263	gi13542685	Mus musculus	SAR1a gene homolog	441	54
1263	gi21634445	Homo sapiens	AF274026_1 GTP-binding protein Sara	446	57
1263	gi33150636	Homo sapiens	AF087897_1 GTP binding protein	446	57
1264	gi22902436	Mus musculus	Sphingosine-1-phosphate phosphatase 1	717	38
1264	gi23345324	Homo sapiens	sphingosine 1-phosphate phosphohydrolase 2	2073	100
1264	gi29436890	Mus musculus	Similar to sphingosine-1-phosphate phosphatase 2	1624	80
1265	gi14	Bos taurus	BoWC1.1	1214	39

239

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1265	gi1480365	Sus scrofa	scavenger-receptor protein	1327	42
1265	gi27464818	Mus musculus	scavenger receptor cysteine-rich type 1 protein CD163c-alpha precursor	1339	44
1266	gi14	Bos taurus	BoWC1.1	1214	39
1266	gi1480365	Sus scrofa	scavenger-receptor protein	1327	42
1266	gi27464818	Mus musculus	scavenger receptor cysteine-rich type 1 protein CD163c-alpha precursor	1339	44
1268	gi21619491	Homo sapiens	similar to expressed sequence AW049604	778	100
1268	gi32967233	Homo sapiens	TAF4A	778	100
1268	gi32967245	Mus musculus	TAF4A	698	93
1270	gi18033185	Danio rerio	AF330001_1 UNC45-related protein	3100	73
1270	gi27436424	Mus musculus	striated muscle UNC45	3937	94
1270	gi27436426	Homo sapiens	striated muscle UNC45	4092	99
1271	gi21064657	Drosophila melanogaster	RH01479p	182	39
1271	gi28375475	Homo sapiens	unnamed protein product	639	99
1271	gi7304173	Drosophila melanogaster	CG1577-PA	182	39
1272	gi16876958	Homo sapiens	AAH16754 hypothetical protein MGC12217	410	100
1273	gi15823642	Homo sapiens	ALS2CR7	2038	100
1273	gi32485022	Homo sapiens	serine/threonine protein kinase	2038	100
1273	gi32485027	Homo sapiens	serine/threonine protein kinase	2320	100
1274	gi12654893	Homo sapiens	AAH01291	400	97
1274	gi2407911	Homo sapiens	CO16	714	96
1274	gi6733554	unidentified	unnamed protein product	710	96
1275	gi18147612	Homo sapiens	metalloprotease disintegrin	4434	95
1275	gi21908028	Homo sapiens	AF466287_1 a disintegrin and metalloprotease domain 33	4434	95
1275	gi21908030	Homo sapiens	a disintegrin and metalloprotease domain 33	4434	95
1276	gi16551401	Homo sapiens	unnamed protein product	2735	100
1276	gi4972116	Arabidopsis thaliana	putative proline-rich protein	133	44
1276	gi7269638	Arabidopsis thaliana	putative proline-rich protein	133	44
1277	gi15291913	Drosophila melanogaster	LD31582p	204	23
1277	gi22477165	Homo sapiens		2783	100
1277	gi26326895	Mus musculus	unnamed protein product	1752	69
1278	gi3452275	Pseudopleuronectes americanus	aminopeptidase N	1008	37
1278	gi525287	Sus scrofa	aminopeptidase N.	1014	38
1278	gi544755	Oryctolagus cuniculus	aminopeptidase N; APN	1021	37
1279	gi13559063	Homo sapiens		747	100
1279	gi24416538	Mus musculus	1700001D09Rik protein	708	71
1279	gi9963863	Homo sapiens	AF226731_1 AD026	738	98
1281	gi20810533	Homo sapiens	hypothetical gene supported by AK054745; AK054745; AK054745; AK054745	414	100
1282	gi20810533	Homo sapiens	hypothetical gene supported by	795	100

240

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			AK054745; AK054745; AK054745; AK054745		
1282	gi26345254	Mus musculus	unnamed protein product	367	63
1282	gi33244011	Mus musculus		374	64
1283	gi20810533	Homo sapiens	hypothetical gene supported by AK054745; AK054745; AK054745; AK054745	789	99
1283	gi26345254	Mus musculus	unnamed protein product	396	64
1283	gi33244011	Mus musculus		403	65
1284	gi18447388	Drosophila melanogaster	RE05944p	700	31
1284	gi21645210	Drosophila melanogaster	CG30394-PA	700	31
1284	gi21645211	Drosophila melanogaster	CG30394-PB	700	31
1285	gi14035874	Homo sapiens	unnamed protein product	910	99
1285	gi14035876	Homo sapiens	unnamed protein product	853	99
1285	gi20070842	Homo sapiens	similar to hypothetical protein FLJ13448	997	99
1286	gi19070822	Mus musculus	AF364868_1 Myb protein P42POP	145	23
1286	gi20977688	Xenopus laevis	tumorhead	146	33
1286	gi27881626	Homo sapiens	LOC339344 protein	150	25
1287	gi10433236	Homo sapiens	unnamed protein product	721	99
1288	gi13278415	Mus musculus	cDNA sequence BC004018	2402	98
1288	gi26355239	Mus musculus	unnamed protein product	2256	97
1288	gi30354720	Mus musculus	AI427653 protein	1357	57
1289	gi12698037	Homo sapiens	KIAA1746 protein	5541	100
1289	gi16769274	Drosophila melanogaster	LD22423p	210	24
1289	gi7298509	Drosophila melanogaster	CG18398-PA	214	24
1290	gi21391484	Homo sapiens	leucine-rich repeat domain- containing protein	397	39
1290	gi21391486	Mus musculus	leucine-rich repeat domain- containing protein	433	40
1290	gi21623740	Rattus norvegicus	Leucine-rich repeat-containing protein 3	428	40
1291	gi20269073	Homo sapiens	putative lipid kinase	2006	76
1291	gi21624340	Homo sapiens	ceramide kinase	2006	76
1291	gi21624342	Mus musculus	ceramide kinases	1617	64
1292	gi312590	Mus musculus	biliary glycoprotein	193	32
1292	gi3549152	Homo sapiens	R29124_1	175	31
1292	gi7414626	Rattus norvegicus	carcinoembryonic antigen- related cell adhesion molecule, secreted isoform CEACAM1a- 4C1	176	31
1293	gi1197500	Homo sapiens	T-cell surface antigen	182	22
1293	gi21707370	Homo sapiens	, sheep red blood cell receptor	182	22
1293	gi312590	Mus musculus	biliary glycoprotein	193	32
1294	gi18676564	Homo sapiens	FLJ00179 protein	993	99
1294	gi21411450	Mus musculus	C230093N12Rik protein	1159	91
1294	gi28839684	Homo sapiens	Similar to expressed sequence AI426465	1242	99

241

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1295	gi27923578	Mus musculus	cerebellin 4 precursor	970	96
1295	gi33416458	Mus musculus	Cerebellin 2 precursor protein	725	73
1295	gi7708438	Homo sapiens		1020	100
1296	gi18490912	Homo sapiens	neurotensin receptor 2	1950	93
1296	gi23138725	Homo sapiens	Similar to neurotensin receptor 2	1984	99
1296	gi3901028	Homo sapiens	neurotensin receptor 2	1955	93
1297	gi15077861	Mus musculus	AF396877_1 bullous pemphigoid antigen 1-e	11308	84
1297	gi179519	Homo sapiens	bullous pemphigoid antigen	10559	98
1297	gi403124	Homo sapiens	bullous pemphigoid antigen	13047	97
1298	gi15077861	Mus musculus	AF396877_1 bullous pemphigoid antigen 1-e	11308	84
1298	gi179519	Homo sapiens	bullous pemphigoid antigen	10559	98
1298	gi403124	Homo sapiens	bullous pemphigoid antigen	13047	97
1299	gi27469519	Homo sapiens	Similar to KIAA0476 gene product	1506	100
1299	gi30268290	Homo sapiens	hypothetical protein	1506	100
1299	gi33330327	Homo sapiens	c-MYC promoter-binding protein IRLB	1501	100
1300	gi15929770	Mus musculus	expressed sequence AW049604	666	100
1300	gi32967235	Homo sapiens	TAF5	666	100
1300	gi32967247	Mus musculus	TAF5	666	100
1301	gi16041156	Macaca fascicularis	X-ray radiation resistance associated 1 protein	729	95
1301	gi18676652	Homo sapiens	FLJ00225 protein	779	100
1301	gi33150874	Homo sapiens	AF439934_1 unknown	779	100
1302	gi16041156	Macaca fascicularis	X-ray radiation resistance associated 1 protein	411	93
1302	gi18676652	Homo sapiens	FLJ00225 protein	444	97
1302	gi33150874	Homo sapiens	AF439934_1 unknown	444	97
1303	gi21619156	Homo sapiens	somatostatin	226	100
1303	gi338288	Homo sapiens	preprosomatostatin I	226	100
1303	gi342299	Macaca fascicularis	preprosomatostatin	226	100
1304	gi22761332	Homo sapiens	unnamed protein product	2052	82
1304	gi24981080	Mus musculus	1810005H09Rik protein	1103	55
1304	gi33417011	Mus musculus		2037	93
1305	gi22761332	Homo sapiens	unnamed protein product	3143	100
1305	gi26331032	Mus musculus	unnamed protein product	2468	81
1305	gi33417011	Mus musculus		2453	85
1306	gi21744725	Homo sapiens	AF478693_1 glycosyl-phosphatidyl-inositol-MAM	1541	48
1306	gi25005320	Sus scrofa	glycosylphosphatidylinositol anchor 1 protein	1536	48
1306	gi33149988	Homo sapiens	MAM domain containing 1	3035	100
1307	gi16550524	Homo sapiens	unnamed protein product	799	100
1308	gi20379980	Mus musculus	2410021P16Rik protein	1731	44
1308	gi22137453	Mus musculus	2410021P16Rik protein	1734	44
1308	gi28280023	Mus musculus	5730439E10Rik protein	3348	80
1309	gi20379980	Mus musculus	2410021P16Rik protein	1634	42
1309	gi22137453	Mus musculus	2410021P16Rik protein	1637	43
1309	gi28280023	Mus musculus	5730439E10Rik protein	3226	78
1310	gi19070124	Mus musculus	AF233346_1 zinc transporter-	1087	95

242

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			like 3 protein		
1310	gi20563194	Mus musculus	AF395840_1 zinc transporter 6	1075	94
1310	gi33338012	Homo sapiens	AF173387_1 MSTP103	942	95
1311	gi12053097	Homo sapiens	hypothetical protein	2127	99
1311	gi23170343	Drosophila melanogaster	CG31556-PA	199	29
1311	gi854065	Human herpesvirus 6	U88	223	32
1312	gi18605758	Mus musculus	9030409G11Rik protein	1343	98
1312	gi6526769	Homo sapiens	HRIHFB2003	1055	97
1312	gi7291408	Drosophila melanogaster	CG11206-PA	822	36
1313	gi19263985	Homo sapiens	Hypothetical protein MGC26766	1565	99
1313	gi19528309	Drosophila melanogaster	LD02310p	573	55
1313	gi7294955	Drosophila melanogaster	CG4080-PA	573	55
1314	gi15030250	Mus musculus	Ureb1-pending protein	5270	95
1314	gi22090626	Homo sapiens	HECT domain protein LASU1	11690	99
1314	gi6841194	Homo sapiens	AF161390_1 HSPC272	9665	99
1315	gi13182757	Homo sapiens	AF212238_1 HTPAP	781	89
1315	gi21542541	Homo sapiens	Similar to HTPAP protein	1074	91
1315	gi28381093	Drosophila melanogaster	CG12746-PD	421	37
1316	gi13182757	Homo sapiens	AF212238_1 HTPAP	915	100
1316	gi21542541	Homo sapiens	Similar to HTPAP protein	1204	99
1316	gi28381093	Drosophila melanogaster	CG12746-PD	539	43
1317	gi14424540	Homo sapiens	AAH09293	1146	93
1317	gi15342051	Homo sapiens	AAH13297	1146	93
1317	gi30582231	Homo sapiens		1146	93
1319	gi14715055	Homo sapiens	MGC9564 protein	487	31
1319	gi16416764	Homo sapiens	AF315594_1 FKSG16	2369	99
1319	gi29436772	Danio rerio	Similar to DNA segment, Chr 11, ERATO Doi 18, expressed	514	30
1320	gi13905212	Mus musculus	RIKEN cDNA 1200006F02	257	77
1320	gi16416764	Homo sapiens	AF315594_1 FKSG16	323	98
1320	gi31873637	Homo sapiens	hypothetical protein	323	98
1321	gi32330803	Mus musculus	podocan protein	2839	91
1321	gi32330805	Homo sapiens	podocan protein	3143	99
1321	gi33636569	Drosophila melanogaster	RE27764p	397	27
1322	gi20258604	Homo sapiens	sialic acid binding Ig-like lectin 5	1470	84
1322	gi20988662	Homo sapiens	sialic acid binding Ig-like lectin 5	1470	84
1322	gi9454520	Homo sapiens	AC018755_5 SIGLEC5	1470	84
1323	gi20258604	Homo sapiens	sialic acid binding Ig-like lectin 5	1372	87
1323	gi20988662	Homo sapiens	sialic acid binding Ig-like lectin 5	1372	87
1323	gi9454520	Homo sapiens	AC018755_5 SIGLEC5	1372	87
1324	gi13183078	Homo sapiens	AF237652_1 a disintegrin-like and metalloprotease domain	602	74

243

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			with thrombospondin type I motifs-like 3		
1324	gi15099921	Homo sapiens	AF176313_1 ADAM-TS related protein 1	874	98
1324	gi20987759	Homo sapiens	Similar to ADAMTS-like 1	886	99
1325	gi178836	Homo sapiens	apolipoprotein C-II	424	89
1325	gi30582255	Homo sapiens	apolipoprotein C-II	418	88
1325	gi757915	Homo sapiens	apoCII protein	424	89
1326	gi178836	Homo sapiens	apolipoprotein C-II	424	89
1326	gi30584853	synthetic construct	Homo sapiens apolipoprotein C-II	422	88
1326	gi757915	Homo sapiens	apoCII protein	424	89
1327	gi15779162	Homo sapiens	AAH14644	477	100
1327	gi21619424	Homo sapiens	Similar to LOC150580	477	100
1328	gi14715231	Homo sapiens	DMBT1/8kb.2 protein	1486	40
1328	gi4105084	Oryctolagus cuniculus	hensin	1428	39
1328	gi6624922	Homo sapiens	DMBT1/8kb.1 protein	1494	41
1329	gi16033591	Homo sapiens	AF416902_1 SH2 domain-containing phosphatase anchor protein 2b	991	99
1329	gi16033597	Homo sapiens	AF416904_1 SH2 domain-containing phosphatase anchor protein 2d	1003	99
1329	gi20810036	Homo sapiens	Fc receptor-like protein 3	985	99
1330	gi28974490	Homo sapiens	lipoma HMGIC fusion-partner-like protein	1183	100
1330	gi30102428	Rattus norvegicus	HMGIC fusion-partner-like protein	1147	95
1330	gi30411045	Mus musculus	Similar to lipoma HMGIC fusion partner	1143	94
1331	gi12060826	Homo sapiens	AF308287_1 serologically defined breast cancer antigen NY-BR-20	607	77
1331	gi17426418	Mus musculus	calmodulin-related protein	788	100
1331	gi19484098	Mus musculus	calmodulin-like 4	783	99
1332	gi10726831	Drosophila melanogaster	CG9986-PA	141	25
1332	gi16741164	Mus musculus	DNA segment, Chr 6, Wayne State University 163, expressed	938	100
1332	gi17862436	Drosophila melanogaster	LD27564p	141	25
1333	gi11693044	Homo sapiens	WNT6 precursor	2000	100
1333	gi13279251	Homo sapiens	AAH04329 Similar to wingless-related MMTV integration site 6	2000	100
1333	gi30583751	Homo sapiens	wingless-type MMTV integration site family, member 6	2000	100
1334	gi19744304	Homo sapiens	AF461760_1 zinc transporter 5	463	94
1334	gi20135611	Homo sapiens	zinc transporter ZnT-5	463	94
1334	gi23270961	Mus musculus	Similar to zinc transporter ZTL1	405	85
1335	gi18480366	Mus musculus	olfactory receptor MOR145-1	310	74
1335	gi21928214	Homo sapiens	seven transmembrane helix	301	77

244

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			receptor		
1335	gi32063318	Mus musculus	olfactory receptor GA_x6K02T2PVTD- 14054886-14053957	310	74
1336	gi12654633	Homo sapiens	Protein inhibitor of activated STAT3	3277	100
1336	gi20988856	Homo sapiens	protein inhibitor of activated STAT3	3277	100
1336	gi30582911	Homo sapiens	protein inhibitor of activated STAT3	3277	100
1337	gi27449075	Oreochromis mossambicus	stearoyl-CoA desaturase	1176	71
1337	gi30350098	Homo sapiens	AF389338_1 acyl-CoA- desaturase	1769	99
1337	gi4469173	Gallus gallus	delta-9 desaturase	1149	71
1338	gi14030861	Homo sapiens	paraneoplastic neuronal antigen MA1	1830	99
1338	gi22726261	Homo sapiens	AF320308_1 paraneoplastic antigen; MA1	1834	100
1338	gi24658774	Homo sapiens	paraneoplastic antigen MA1	1834	100
1339	gi29468118	Homo sapiens	AF357888_1 PAP-2-like protein 2	1695	100
1339	gi31580553	Homo sapiens	plasticity related gene 2	1695	100
1339	gi32186953	Homo sapiens	lipid phosphate phosphatase- related protein type 3	1695	100
1340	gi11137605	Homo sapiens		1931	100
1340	gi20809333	Homo sapiens	actin like protein	1928	99
1340	gi684936	Homo sapiens	peptide with resemblance to the actin family; the actual start of the coding region has not been determined	1362	88
1341	gi11177510	Rattus norvegicus	AF287300_1 tandem pore domain potassium channel THIK-2	2215	98
1341	gi11177514	Homo sapiens	AF287302_1 tandem pore domain potassium channel THIK-2	2234	100
1341	gi28839529	Homo sapiens	Potassium channel, subfamily K, member 12	2234	100
1342	gi14198194	Mus musculus	CDNA sequence BC008155	606	77
1342	gi14336716	Homo sapiens	AE006464_16 similar to FBan0003337	1216	100
1342	gi7300722	Drosophila melanogaster	CG3337-PA	326	40
1343	gi11862939	Mus musculus	DDM36	1117	43
1343	gi11862941	Mus musculus	DDM36E	1105	43
1343	gi19570398	Homo sapiens	hDDM36	1120	43
1344	gi21744725	Homo sapiens	AF478693_1 glycosyl- phosphatidyl-inositol-MAM	4898	98
1344	gi25005318	Sus scrofa	MAM domain containing glycosylphosphatidylinositol anchor 1	4355	95
1344	gi25005320	Sus scrofa	glycosylphosphatidylinositol anchor 1 protein	4224	94
1345	gi12276198	Homo sapiens	AF333487_1 FKSG40	1020	100

245

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1345	gi12408250	Homo sapiens	FKSG28	1020	100
1345	gi18652934	Xenopus laevis	Mig30	634	49
1346	gi21410151	Mus musculus	LOC213895 protein	1657	73
1346	gi27696627	Homo sapiens	Ribosome biogenesis protein BMS1 homolog	4190	99
1346	gi7294027	Drosophila melanogaster	CG7728-PA	1345	43
1347	gi12842044	Mus musculus	unnamed protein product	554	71
1347	gi18921437	Mus musculus	2010004A03Rik protein	850	70
1347	gi20987450	Homo sapiens	LOC146433	1160	95
1348	gi1016012	Rattus norvegicus	neural cell adhesion protein BIG-2 precursor	5147	92
1348	gi26891535	Homo sapiens	contactin 4	5366	98
1348	gi29837411	Homo sapiens	BIG-2	5366	98
1349	gi30102449	Homo sapiens	lipoma HMGIC fusion-partner-like protein	1161	97
1349	gi30908798	Homo sapiens	lipoma HMGIC fusion partner-like protein 4	952	80
1349	gi30908800	Rattus norvegicus	lipoma HMGIC fusion partner-like protein 4	951	80
1350	gi13097705	Homo sapiens	AAH03559 , member 3	2028	95
1350	gi1340142	Homo sapiens	alpha1-antichymotrypsin	2024	95
1350	gi21961493	Homo sapiens	, member 3	2025	95
1351	gi1850850	Murid herpesvirus 4	serine threonine rich glycoprotein	166	30
1351	gi21618556	Homo sapiens		3529	91
1351	gi33304372	Homo sapiens	tastin	3524	91
1352	gi12053849	Homo sapiens	DREV protein	1689	100
1352	gi12053851	Homo sapiens	DREV1 protein	1673	99
1352	gi12053853	Homo sapiens	DREV protein	1689	100
1353	gi14627081	Homo sapiens	AF367017_1 caspase-1 dominant-negative inhibitor Pseudo-ICE	492	100
1353	gi21707335	Homo sapiens	Similar to CARD only protein	462	100
1353	gi33793	Homo sapiens	interleukin-1B converting enzyme	445	92
1355	gi22760096	Homo sapiens	unnamed protein product	1051	93
1355	gi27883913	Homo sapiens	POTE	497	48
1355	gi28279813	Homo sapiens	Similar to hypothetical protein DKFZp434A171	860	99
1356	gi11125348	Homo sapiens	putative protein kinase	11920	99
1356	gi6933864	Homo sapiens	kinase deficient protein KDP	3408	100
1356	gi8272557	Rattus norvegicus	AF227741_1 protein kinase WNK1	5436	73
1357	gi11125348	Homo sapiens	putative protein kinase	9671	99
1357	gi20987908	Mus musculus	LOC269796 protein	1553	82
1357	gi8272557	Rattus norvegicus	AF227741_1 protein kinase WNK1	5436	73
1358	gi10946203	Homo sapiens	AF272363_1 neuromedin U receptor 2	785	100
1358	gi16877377	Homo sapiens	AAH16938 neuromedin U receptor 2	785	100
1358	gi9944990	Homo sapiens	AF292402_1 neuromedin U receptor-type 2	785	100

246

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1359	gi15020809	Takifugu rubripes	putative methionyl tRNA synthetase	1823	64
1359	gi17861592	Drosophila melanogaster	GH13807p	1212	45
1359	gi23171238	Drosophila melanogaster	CG31322-PA	1212	45
1360	gi15341975	Homo sapiens	AAH13184 Similar to major histocompatibility complex, class II, DP beta 1	437	72
1360	gi17389919	Homo sapiens	AAH17967 Similar to major histocompatibility complex, class II, DP beta 1	819	100
1360	gi188479	Homo sapiens	HLA-DPB1	437	72
1361	gi19701013	Homo sapiens	unnamed protein product	1143	99
1361	gi3342737	Homo sapiens	R26660_2, partial CDS	1024	100
1361	gi3478640	Homo sapiens	R26660_2, partial CDS	154	100
1362	gi15779083	Homo sapiens	AAH14609	1172	99
1362	gi3342737	Homo sapiens	R26660_2, partial CDS	1002	96
1362	gi3478640	Homo sapiens	R26660_2, partial CDS	154	100
1363	gi13991167	Homo sapiens	sialic acid-binding immunoglobulin-like lectin-like long splice variant	2879	99
1363	gi14625822	Homo sapiens	AF282256_1 Siglec-L1	2879	99
1363	gi23272769	Homo sapiens	SIGLEC-like 1	2879	99
1364	gi15132186	Homo sapiens	unnamed protein product	1644	100
1364	gi15132529	Homo sapiens	unnamed protein product	1644	100
1364	gi21439502	Homo sapiens	unnamed protein product	1644	100
1365	gi19353230	Homo sapiens	interleukin 1, delta	823	100
1365	gi6165336	Homo sapiens	interleukin-1-like protein 1	823	100
1365	gi9651789	Homo sapiens	AF230377_1 interleukin-1 delta	823	100
1366	gi177870	Homo sapiens	alpha-2-macroglobulin precursor	2765	40
1366	gi25303946	Homo sapiens	alpha-2-macroglobulin	2765	40
1366	gi579594	Homo sapiens	alpha 2-macroglobulin 690-740	2760	40
1367	gi25990364	Homo sapiens	AF319622_1 P-glycoprotein	555	98
1367	gi27656757	Takifugu rubripes	Mdr3	311	52
1367	gi4574224	Fundulus heteroclitus	AF099732_1 multidrug resistance transporter homolog	287	49
1368	gi12805221	Mus musculus	Lymphocyte antigen 6 complex, locus A	713	100
1368	gi198924	Mus musculus	Ly-6A.2	713	100
1368	gi201113	Mus musculus	T-cell activation protein	713	100
1967	gi13543526	Homo sapiens	AAH05921	616	96
1967	gi18088830	Homo sapiens	AAH20756	616	96
1967	gi30582691	Homo sapiens		616	96
1968	gi13543526	Homo sapiens	AAH05921	616	96
1968	gi18088830	Homo sapiens	AAH20756	616	96
1968	gi30582691	Homo sapiens		616	96
1969	gi13543526	Homo sapiens	AAH05921	616	96
1969	gi18088830	Homo sapiens	AAH20756	616	96
1969	gi30582691	Homo sapiens		616	96
1970	gi13543526	Homo sapiens	AAH05921	616	96
1970	gi18088830	Homo sapiens	AAH20756	616	96

247

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1970	gi30582691	Homo sapiens		616	96
1971	gi12653501	Homo sapiens	SERPINF1 protein	2119	99
1971	gi15217079	Homo sapiens	AF400442_1 pigment epithelium-derived factor	2125	99
1971	gi30583283	Homo sapiens	, member 1	2119	99
1972	gi20269957	Sus scrofa	AF498759_1 phospholipase C delta 4	166	96
1972	gi21307610	Mus musculus	phospholipase C delta 4	158	90
1972	gi571466	Rattus norvegicus	phospholipase C delta-4	151	84
1973	gi17864023	Homo sapiens	AF450090_1 KCCR13L	3299	94
1973	gi22760385	Homo sapiens	unnamed protein product	3290	94
1973	gi22761016	Homo sapiens	unnamed protein product	3299	94
1975	gi19684107	Homo sapiens		120	92
1975	gi32966069	Homo sapiens	CD39L2 nucleotidase	120	92
1975	gi4691263	Homo sapiens		120	92
1976	gi11493483	Homo sapiens	AF130117_48 PRO2550	364	71
1976	gi2580578	Homo sapiens	ubiquitous TPR motif, Y isoform	339	75
1976	gi8572229	Homo sapiens	ubiquitous TPR-motif protein Y isoform	339	75
1977	gi18848355	Mus musculus	Coq6 protein	2085	87
1977	gi30047245	Mus musculus	Coq6 protein	2090	85
1977	gi4680659	Homo sapiens	AF132944_1 CGI-10 protein	2378	98
1978	gi12654881	Homo sapiens	AAH01284	331	78
1978	gi1710216	Homo sapiens	unknown	311	73
1978	gi28799226	Homo sapiens	unnamed protein product	252	65
1979	gi11493483	Homo sapiens	AF130117_48 PRO2550	143	48
1979	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	161	63
1979	gi32486167	Homo sapiens	AD7C-NTP	161	63
1980	gi20810589	Homo sapiens	similar to arsenite inducible RNA associated protein	833	99
1980	gi22945274	Drosophila melanogaster	CG12795-PA	455	54
1980	gi9651711	Mus musculus	AF224494_1 arsenite inducible RNA associated protein	687	80
1981	gi13241652	Rattus norvegicus	AF309558_1 supernatant protein factor	162	87
1981	gi13543184	Mus musculus	SEC14-like 2	162	87
1981	gi6624130	Rattus norvegicus	AC004832_1 similar to 45 kDa secretory protein	169	96
1982	gi11066250	Homo sapiens	AF197937_1 presenilins associated rhomboid-like protein	1392	100
1982	gi13177766	Homo sapiens	AAH03653 Similar to presenilins associated rhomboid-like protein	1068	80
1982	gi15559382	Homo sapiens	AAH14058 presenilins associated rhomboid-like protein	1389	99
1983	gi1864091	Rattus norvegicus	PSD-95/SAP90-associated protein-3	160	100
1984	gi11877274	Homo sapiens		2265	100
1984	gi21667210	Homo sapiens	AF465765_1	2265	100

248

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			bactericidal/permeability-increasing protein-like 1		
1984	gi21706776	Homo sapiens	Bactericidal/permeability-increasing protein-like 1	2258	99
1985	gi3879547	Caenorhabditis elegans		125	36
1986	gi21307771	Homo sapiens	organic anion transporter 2	733	100
1986	gi21707474	Homo sapiens	, member 7	733	100
1986	gi5001689	Homo sapiens	AF097518_1 liver-specific transporter	733	100
1987	gi12804105	Homo sapiens	AAH02905 Similar to CG15084 gene product	589	79
1987	gi13649459	Homo sapiens	AF250306_1 putative SB115 protein	589	79
1987	gi18204670	Mus musculus	4930527D15Rik protein	569	75
1988	gi1022323	Mus musculus	chain	3354	87
1988	gi537329	Homo sapiens	alpha-2 type IV collagen	3752	99
1988	gi556299	Mus musculus	alpha-2 type IV collagen	3351	87
1989	gi17298315	Homo sapiens	candidate tumor suppressor protein	1360	98
1989	gi7861733	Homo sapiens	AF176832_1 low density lipoprotein receptor related protein-deleted in tumor	1360	98
1989	gi8926243	Mus musculus	AF270884_1 low density lipoprotein receptor related protein LRP1B/LRP-DIT	1181	84
1990	gi17298315	Homo sapiens	candidate tumor suppressor protein	1360	98
1990	gi7861733	Homo sapiens	AF176832_1 low density lipoprotein receptor related protein-deleted in tumor	1360	98
1990	gi8926243	Mus musculus	AF270884_1 low density lipoprotein receptor related protein LRP1B/LRP-DIT	1181	84
1991	gi11493483	Homo sapiens	AF130117_48 PRO2550	408	78
1991	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	328	75
1991	gi7770139	Homo sapiens	AF119917_13 PRO1722	328	72
1992	gi157409	Drosophila melanogaster	fat protein	370	37
1992	gi23093109	Drosophila melanogaster	CG7749-PA	367	41
1992	gi7295732	Drosophila melanogaster	CG3352-PA	367	38
1993	gi157409	Drosophila melanogaster	fat protein	370	37
1993	gi23093109	Drosophila melanogaster	CG7749-PA	367	41
1993	gi7295732	Drosophila melanogaster	CG3352-PA	367	38
1994	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	410	89
1994	gi29293087	Homo sapiens	dipeptidyl peptidase 9	410	89
1994	gi3513303	Homo sapiens	R26984_1	476	100
1995	gi32493172	Homo sapiens	pheromone receptor	170	96

249
TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
1995	gi32493174	Homo sapiens	pheromone receptor	170	96
1995	gi32493176	Homo sapiens	pheromone receptor	178	100
1996	gi23468368	Mus musculus	1200013F24Rik protein	799	63
1996	gi27695305	Mus musculus	1200013F24Rik protein	825	76
1996	gi7582294	Homo sapiens	AF208853_1 BM-011	781	98
1997	gi1620870	Ciona intestinalis	myoplasmin-C1	190	29
1997	gi31419817	Mus musculus	Golgi autoantigen, golgin subfamily a, 3	124	26
1997	gi4582571	Gallus gallus	Hyperion protein, 419 kD isoform	125	26
1998	gi13872813	Homo sapiens	fibulin-6	1099	48
1998	gi14575679	Homo sapiens	AF156100_1 hemicentin	2159	86
1998	gi3879658	Caenorhabditis elegans		636	32
1999	gi14044052	Homo sapiens	AAH07950	1105	51
1999	gi17390825	Mus musculus	heterogenous nuclear ribonucleoprotein U	1104	51
1999	gi3822553	Gallus gallus	nuclear calmodulin-binding protein	1554	64
2000	gi17223626	Homo sapiens	ATP-binding cassette A10	1683	93
2000	gi32350914	Homo sapiens	ATP-binding cassette sub-family A member 10	1675	92
2000	gi32350969	Homo sapiens	ATP-binding cassette sub-family A member 10	1675	92
2001	gi13374079	Homo sapiens	TAFII140 protein	3747	99
2001	gi13374178	Mus musculus	TAFII140 protein	3454	85
2001	gi28175603	Homo sapiens	TAF3 protein	2775	99
2002	gi17429038	Ralstonia solanacearum	PROBABLE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE PROTEIN	676	61
2002	gi22776354	Oceanobacillus ihayensis HTE831	acyl-CoA dehydrogenase	660	63
2002	gi28280023	Mus musculus	5730439E10Rik protein	974	84
2003	gi21522776	Homo sapiens	unnamed protein product	2998	98
2003	gi24047224	Homo sapiens	Similar to EGF-like-domain, multiple 6	2982	98
2003	gi6752658	Homo sapiens	AF186084_1 epidermal growth factor repeat containing protein	2984	98
2004	gi14530342	Caenorhabditis elegans		389	51
2004	gi6531661	Caenorhabditis elegans	AF195610_1 LIN-41A	389	51
2004	gi6531663	Caenorhabditis elegans	AF195611_1 LIN-41B	389	51
2005	gi1504026	Homo sapiens		5996	99
2005	gi22725157	Homo sapiens	minor histocompatibility antigen HA-1	5835	99
2005	gi23272016	Homo sapiens	Similar to PTPL1-associated RhoGAP 1	5675	98
2006	gi13274120	Homo sapiens		995	91
2006	gi6102996	Mus musculus	Vanin-3	884	78
2006	gi7160973	Homo sapiens	VNN3 protein	995	91
2007	gi27463365	Homo sapiens	a disintegrin-like and metalloprotease with	345	93

250
TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			thrombospondin type 1 motifs 9B		
2007	gi3876367	Caenorhabditis elegans		148	39
2007	gi3879882	Caenorhabditis elegans		148	39
2008	gi15963476	Homo sapiens	AF289221_1 alpha-adaptin A related protein	2085	94
2008	gi15963477	Homo sapiens	AF289221_2 alpha-adaptin A related protein	2118	99
2008	gi4314340	AA 159-977	Human alpha-adaptin A homolog	2085	94
2009	gi15488017	Homo sapiens	AF407274_1 EW12	3200	100
2009	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	3200	100
2009	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	3200	100
2010	gi15488017	Homo sapiens	AF407274_1 EW12	3200	100
2010	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	3200	100
2010	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	3200	100
2011	gi15488017	Homo sapiens	AF407274_1 EW12	3200	100
2011	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	3200	100
2011	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	3200	100
2012	gi15488017	Homo sapiens	AF407274_1 EW12	3200	100
2012	gi27497567	Homo sapiens	keratinocytes associated transmembrane protein 4	3200	100
2012	gi31753233	Homo sapiens	Immunoglobulin superfamily, member 8	3200	100
2013	gi1405723	Homo sapiens	type X collagen	198	30
2013	gi30095	Homo sapiens	3	198	30
2013	gi7573532	Homo sapiens		198	30
2014	gi15145793	Sus scrofa	basic proline-rich protein	233	26
2014	gi15145795	Sus scrofa	basic proline-rich protein	205	26
2014	gi25056007	Zea mays	AF159297_1 extensin-like protein	203	26
2015	gi21992	Volvox carteri	extensin	158	37
2015	gi2429362	Santalum album	proline rich protein	166	39
2015	gi32488576	Oryza sativa (japonica cultivar-group)	OSJNBa0067K08.27	157	35
2016	gi12002042	Homo sapiens	AF063606_1 brain my048 protein	659	70
2016	gi17225331	Homo sapiens	AF325115_1 MY0876G05 protein	659	70
2016	gi17646146	Homo sapiens	AF314542_1 B lymphocyte activation-related protein	727	56
2018	gi13161063	Homo sapiens	AF332218_1 protocadherin 11	746	56
2018	gi13161066	Homo sapiens	AF332219_1 protocadherin 11	746	56
2018	gi9845485	Homo sapiens	AF169692_1 protocadherin-9	1349	100
2019	gi16552038	Homo sapiens	unnamed protein product	2139	99
2019	gi21410124	Mus musculus	3230402E02Rik protein	1334	60

251

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2019	gi5688958	Homo sapiens	PMMLP	2140	100
2020	gi21734445	Rattus norvegicus	BMP/Retinoic acid-inducible neurai-specific protein-2	3958	95
2020	gi21734447	Rattus norvegicus	BMP/Retinoic acid-inducible neural-specific protein-3	2948	70
2020	gi30348610	Gallus gallus	BMP/retinoic acid-inducible neural-specific protein	2090	52
2021	gi23272677	Homo sapiens	Similar to zinc finger protein 208	467	80
2021	gi26251755	Homo sapiens	ZNF431 protein	449	78
2021	gi30421228	Homo sapiens	zinc finger protein 430	572	100
2022	gi23272677	Homo sapiens	Similar to zinc finger protein 208	467	80
2022	gi26251755	Homo sapiens	ZNF431 protein	449	78
2022	gi30421228	Homo sapiens	zinc finger protein 430	572	100
2023	gi1212965	Homo sapiens	transmembrane protein	358	70
2023	gi1213221	Rattus norvegicus	transmembrane protein	354	69
2023	gi19683999	Homo sapiens	coated vesicle membrane protein	358	70
2024	gi1199524	Homo sapiens	acid phosphatase	2246	99
2024	gi13111975	Homo sapiens	AAH03160 acid phosphatase 2, lysosomal	2242	99
2024	gi30584617	synthetic construct	Homo sapiens acid phosphatase 2, lysosomal	2242	99
2025	gi15625570	Homo sapiens	AF411981_1 centaurin beta5	353	100
2025	gi30109272	Homo sapiens	CENTB5 protein	505	99
2025	gi4688902	Homo sapiens	centaurin beta2	270	48
2026	gi27693942	Homo sapiens	Similar to expressed sequence AI449432	1083	42
2026	gi2789430	Homo sapiens	repressor protein	1084	42
2026	gi5630080	Homo sapiens	AC004890_2	1077	42
2027	gi11345382	Homo sapiens	AF308801_1 vacuolar protein sorting protein 16	2977	99
2027	gi12140290	Homo sapiens		2983	99
2027	gi15553046	Mus musculus	Vps16	2932	97
2028	gi30141048	Homo sapiens	Nogo-66 receptor homolog-1	294	100
2028	gi30141052	Rattus norvegicus	Nogo-66 receptor homolog-1	270	92
2028	gi32351287	Rattus norvegicus	Nogo-66 receptor homolog 2	149	53
2029	gi202592	Rattus norvegicus	prealpha-2-macroglobulin	238	40
2029	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2029	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2030	gi15778556	Homo sapiens	AF414429_1 alpha-1-B glycoprotein precursor	131	92
2031	gi200057	Mus musculus	neuronal glycoprotein	698	94
2031	gi29837411	Homo sapiens	BIG-2	554	75
2031	gi563133	Rattus norvegicus	BIG-1 protein	692	94
2032	gi16550078	Homo sapiens	unnamed protein product	763	100
2032	gi28175743	Homo sapiens	similar to hypothetical protein FLJ30803	763	100
2032	gi30354720	Mus musculus	AI427653 protein	756	100
2033	gi16550078	Homo sapiens	unnamed protein product	763	100
2033	gi28175743	Homo sapiens	similar to hypothetical protein FLJ30803	763	100
2033	gi30354720	Mus musculus	AI427653 protein	756	100

252

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2034	gi21929093	Homo sapiens	seven transmembrane helix receptor	1711	88
2034	gi24286029	Homo sapiens	G-protein coupled receptor GPR116	6754	97
2034	gi5525078	Rattus norvegicus	seven transmembrane receptor	5038	72
2035	gi11917507	Homo sapiens	HPF1 protein	434	59
2035	gi13938351	Homo sapiens	AAH07307 Similar to zinc finger protein 268	432	63
2035	gi3135968	Homo sapiens		440	58
2036	gi13097633	Homo sapiens	AAH03534 Similar to ATPase, Class I, type 8B, member 1	373	84
2036	gi33440008	Homo sapiens	possible aminophospholipid translocase ATP8B2	406	91
2036	gi3628757	Homo sapiens	FIC1	373	84
2038	gi11558486	Homo sapiens	B-cell lymphoma/leukaemia 11A short form	1314	99
2038	gi12150278	Homo sapiens	AF080216_1 C2H2-type zinc-finger protein; EVI-9	1197	98
2038	gi30410854	Mus musculus		1312	98
2039	gi32394378	Homo sapiens	forkhead-associated domain histidine-triad like protein	1735	94
2039	gi32394380	Bos taurus	forkhead-associated domain histidine-triad like protein	1540	83
2039	gi32394382	Sus scrofa	forkhead-associated domain histidine-triad like protein	1575	84
2040	gi32394378	Homo sapiens	forkhead-associated domain histidine-triad like protein	1735	94
2040	gi32394380	Bos taurus	forkhead-associated domain histidine-triad like protein	1540	83
2040	gi32394382	Sus scrofa	forkhead-associated domain histidine-triad like protein	1575	84
2041	gi32394378	Homo sapiens	forkhead-associated domain histidine-triad like protein	1735	94
2041	gi32394380	Bos taurus	forkhead-associated domain histidine-triad like protein	1540	83
2041	gi32394382	Sus scrofa	forkhead-associated domain histidine-triad like protein	1575	84
2042	gi26454883	Homo sapiens	hypothetical protein HSPC148	1181	100
2042	gi6523797	Homo sapiens	AF110775_1 adrenal gland protein AD-002	1181	100
2042	gi6841518	Homo sapiens	AF161497_1 HSPC148	1178	99
2043	gi14009597	Homo sapiens	AF282619_1 lysyl oxidase-like 3 protein	1569	98
2043	gi14486600	Homo sapiens	AF311313_1 lysyl oxidase-like 3 protein	1569	98
2043	gi15186770	Homo sapiens	AF284815_1 lysyl oxidase-like protein	1569	98
2044	gi10834722	Homo sapiens	AF258588_1 PP5656	892	89
2044	gi21706836	Mus musculus	Gylt11b protein	1056	87
2044	gi22713410	Homo sapiens	GYLTL1B protein	1205	100
2045	gi7209721	Mus musculus	DD57	2242	88
2045	gi7209723	Homo sapiens	WD-repeat like sequence	2483	100
2045	gi8217485	Homo sapiens		2480	99
2046	gi13592175	Leishmania major	AC084329_1 ppg3	140	28
2046	gi28828184	Dictyostelium	similar to Leishmania major.	179	28

253

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
		discoideum	Ppg3		
2046	gi3873550	Schizosaccharomyces pombe	SPBC215.13	147	24
2047	gi21104460	Homo sapiens	OK/SW-CL.19	206	100
2047	gi32425794	Homo sapiens	NJMU-R1 protein	206	100
2047	gi32450708	Homo sapiens	NJMU-R1 protein	206	100
2048	gi13277972	Mus musculus	phosphatidate cytidyltransferase 2	2270	95
2048	gi19344052	Homo sapiens	...	2360	99
2048	gi4186023	Homo sapiens	CDS2 protein	2360	99
2049	gi17862928	Drosophila melanogaster	SD03549p	121	35
2049	gi29387317	Mus musculus	1200011O22Rik protein	670	89
2049	gi7297878	Drosophila melanogaster	CG14941-PA	121	35
2050	gi13562004	Nephila madagascariensis	AF350276_1 major ampullate spidroin 2-like protein	251	33
2050	gi7106224	Nephila clavipes	flagelliform silk protein	252	32
2050	gi7106228	Nephila inaurata madagascariensis	flagelliform silk protein [Nephila madagascariensis]	277	34
2051	gi12018147	Chlamydomonas reinhardtii	AF309494_1 vegetative cell wall protein gp1	198	31
2051	gi15145793	Sus scrofa	basic proline-rich protein	204	29
2051	gi15145797	Sus scrofa	basic proline-rich protein	200	30
2052	gi16877193	Homo sapiens	AAH16860 G protein-coupled receptor, family C, group 5, member C	2320	99
2052	gi30583709	Homo sapiens	G protein-coupled receptor, family C, group 5, member C	2320	99
2052	gi8118032	Homo sapiens	AF207989_1 orphan G-protein coupled receptor	2320	99
2053	gi15679980	Homo sapiens	C114 protein	930	99
2053	gi16769562	Drosophila melanogaster	LD38910p	328	47
2053	gi7302978	Drosophila melanogaster	CG8441-PA	328	47
2054	gi10726751	Drosophila melanogaster	CG13623-PA	333	53
2054	gi21430012	Drosophila melanogaster	GH27470p	333	53
2054	gi7406400	Arabidopsis thaliana	putative protein	317	45
2055	gi13959018	Homo sapiens	AF361746_1 endothelial cell-selective adhesion molecule	1578	99
2055	gi13991773	Mus musculus	AF361882_1 endothelial cell-selective adhesion molecule	1188	76
2055	gi29165726	Mus musculus	Endothelial cell-selective adhesion molecule	1188	76
2056	gi15422171	Homo sapiens	22 kDa peroxisomal membrane protein 2	862	99
2056	gi297437	Rattus norvegicus	peroxisomal membrane protein	680	76
2056	gi8164184	Homo sapiens	22kDa peroxisomal membrane protein-like	862	99
2057	gi11994465	Arabidopsis thaliana	contains similarity to late embryogenesis abundant protein~gene_id:MLD14.16	141	39

254

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2057	gi21326031	Oryzias latipes	choriogenin H	159	35
2057	gi22093906	Oryzias latipes	AF396668_1 choriogenin H	157	35
2058	gi62877	Gallus gallus	type VI collagen alpha-2 subunit preprotein	320	42
2058	gi62881	Gallus gallus	type VI collagen subunit alpha2	320	42
2058	gi62882	Gallus gallus	type VI collagen subunit alpha2	320	42
2059	gi17945608	Drosophila melanogaster	RE26969p	600	60
2059	gi7292879	Drosophila melanogaster	CG1998-PA	600	60
2059	gi7292910	Drosophila melanogaster	CG11162-PA	423	50
2060	gi17066106	Homo sapiens	Novex-3 Titin Isoform	964	99
2060	gi27696390	Xenopus laevis	Similar to titin	251	37
2060	gi992994	Gallus gallus	myosin light chain kinase	228	35
2061	gi14089982	Mycoplasma pulmonis		143	33
2061	gi2649941	Archaeoglobus fulgidus DSM 4304		151	30
2061	gi30180922	Nitrosomonas europaea ATCC 19718	Adenylate kinase	143	28
2062	gi29477024	Mus musculus	Similar to RIKEN cDNA 9130023G24 gene	464	44
2062	gi3002588	Mus musculus	Plenty of SH3s; POSH	148	25
2062	gi7453547	Homo sapiens	glioma tumor suppressor candidate region protein 1	125	25
2063	gi29477024	Mus musculus	Similar to RIKEN cDNA 9130023G24 gene	464	44
2063	gi3002588	Mus musculus	Plenty of SH3s; POSH	148	25
2063	gi7453547	Homo sapiens	glioma tumor suppressor candidate region protein 1	125	25
2064	gi10441350	Mus musculus	olfactory UDP glucuronosyltransferase	241	70
2064	gi4580602	Macaca fascicularis	AF112112_1 UDP-glucuronosyltransferase 2B19 precursor	244	73
2064	gi4753766	Homo sapiens	UDP glucuronosyltransferase	266	76
2065	gi13325266	Homo sapiens	AAH04450 hypothetical protein MGC2650	796	91
2065	gi3688090	Homo sapiens	R32611_2	827	100
2065	gi6841228	Homo sapiens	AF161407_1 HSPC289	703	84
2066	gi11493483	Homo sapiens	AF130117_48 PRO2550	282	56
2066	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	497	62
2066	gi32486167	Homo sapiens	AD7C-NTP	497	62
2067	gi16552274	Homo sapiens	unnamed protein product	276	45
2067	gi57516	Rattus rattus	ASM15	437	57
2067	gi7107346	Peromyscus maniculatus bairdii	H19	280	43
2068	gi20330550	Homo sapiens	AF251706_1 NK inhibitory receptor precursor	1480	94
2068	gi30962591	Homo sapiens	AF375480_1 immune receptor	1401	93

255

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			expressed on myeloid cells splice variant 1		
2068	gi31790204	Homo sapiens	inhibitory receptor IREM1	1478	94
2069	gi20330550	Homo sapiens	AF251706_1 NK inhibitory receptor precursor	1480	94
2069	gi30962591	Homo sapiens	AF375480_1 immune receptor expressed on myeloid cells splice variant 1	1401	93
2069	gi31790204	Homo sapiens	inhibitory receptor IREM1	1478	94
2070	gi20330550	Homo sapiens	AF251706_1 NK inhibitory receptor precursor	1480	94
2070	gi30962591	Homo sapiens	AF375480_1 immune receptor expressed on myeloid cells splice variant 1	1401	93
2070	gi31790204	Homo sapiens	inhibitory receptor IREM1	1478	94
2071	gi18307481	Homo sapiens	phosphoinositide-binding proteins	2206	97
2071	gi27695704	Mus musculus	Connector enhancer of KSR2	705	35
2071	gi29691916	Rattus norvegicus	interactor protein for cytohesin exchange factors 1	1651	79
2072	gi11493982	Homo sapiens	AF208232_1 TLH29 protein precursor	303	70
2072	gi15929988	Homo sapiens	AAH15423 Similar to TLH29 protein precursor	497	100
2072	gi21618549	Homo sapiens	TLH29 protein precursor	303	70
2073	gi11493982	Homo sapiens	AF208232_1 TLH29 protein precursor	303	70
2073	gi15929988	Homo sapiens	AAH15423 Similar to TLH29 protein precursor	497	100
2073	gi21618549	Homo sapiens	TLH29 protein precursor	303	70
2074	gi12804693	Homo sapiens	AAH01773 Similar to ribosomal protein L34	591	100
2074	gi17932958	Homo sapiens	ribosomal protein L34	591	100
2074	gi20306434	Mus musculus	1100001I22Rik protein	587	99
2075	gi15384841	Homo sapiens	activating NK receptor	738	99
2075	gi15384843	Homo sapiens	NTB-A receptor	754	100
2075	gi20988099	Mus musculus	lymphocyte antigen 108	240	39
2076	gi10177621	Arabidopsis thaliana	phytoene dehydrogenase-like	573	42
2076	gi17979255	Arabidopsis thaliana	AT5g49550/K6M13_10	589	42
2076	gi29028742	Arabidopsis thaliana	At5g49550/K6M13_10	589	42
2077	gi14270364	Mus musculus	Epigen protein	378	71
2077	gi6272269	Rattus norvegicus	NC1 protein	122	52
2077	gi7799191	Mus musculus	tomoregulin-1	122	52
2078	gi14270364	Mus musculus	Epigen protein	378	71
2078	gi6272269	Rattus norvegicus	NC1 protein	122	52
2078	gi7799191	Mus musculus	tomoregulin-1	122	52
2079	gi14270364	Mus musculus	Epigen protein	378	71
2079	gi6272269	Rattus norvegicus	NC1 protein	122	52
2079	gi7799191	Mus musculus	tomoregulin-1	122	52
2080	gi27469556	Homo sapiens	Putative neuronal cell adhesion molecule	206	34
2080	gi29289929	Danio rerio	neogenin	176	37
2080	gi3068592	Mus musculus	punc	192	35
2081	gi31753150	Homo sapiens	Ras family member Ris	665	65

256
TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2081	gi4416181	Mus musculus	ES18	1276	84
2081	gi7331127	Homo sapiens	AF233588_1 Ris	665	65
2082	gi13128925	Homo sapiens	AF304378_1 ULBP2 protein	1312	99
2082	gi18650584	Homo sapiens	retinoic acid early transcript 1	1312	99
2082	gi21961213	Homo sapiens	UL16 binding protein 2	1312	99
2083	gi13872813	Homo sapiens	fibulin-6	513	29
2083	gi14575679	Homo sapiens	AF156100_1 hemicentin	513	29
2083	gi9280405	Homo sapiens	AF245505_1 adlcan	1462	46
2084	gi13872813	Homo sapiens	fibulin-6	513	29
2084	gi14575679	Homo sapiens	AF156100_1 hemicentin	513	29
2084	gi9280405	Homo sapiens	AF245505_1 adlcan	1462	46
2085	gi13872813	Homo sapiens	fibulin-6	513	29
2085	gi14575679	Homo sapiens	AF156100_1 hemicentin	513	29
2085	gi9280405	Homo sapiens	AF245505_1 adlcan	1462	46
2086	gi3041867	Homo sapiens	p53	162	96
2086	gi4731632	Homo sapiens	AF135121_1 tumor suppressor protein p53	162	96
2086	gi4732147	Homo sapiens	AF136271_1 tumor suppressor protein p53	162	96
2087	gi12240284	Mus musculus	AF327059_1 apolipoprotein A5	1300	72
2087	gi6707433	Homo sapiens	AF202889_1 apolipoprotein A5	1864	100
2087	gi6707435	Homo sapiens	AF202890_1 apolipoprotein A5	1864	100
2088	gi12240284	Mus musculus	AF327059_1 apolipoprotein A5	1300	72
2088	gi6707433	Homo sapiens	AF202889_1 apolipoprotein A5	1864	100
2088	gi6707435	Homo sapiens	AF202890_1 apolipoprotein A5	1864	100
2089	gi13111784	Homo sapiens	AAH03081 hypothetical protein FLJ10637	1509	99
2089	gi13543037	Mus musculus	4933424B01Rik protein	958	80
2089	gi14249965	Homo sapiens	AAH08368 hypothetical protein FLJ10637	1513	100
2090	gi19344001	Homo sapiens	phospholipase A2, group IID	846	99
2090	gi5771420	Homo sapiens	AF112982_1 group IID secretory phospholipase A2	852	100
2090	gi6453793	Homo sapiens	AF188625_1 phospholipase A2	846	99
2091	gi1674069	Mycoplasma pneumoniae	30K adhesin-related protein	132	35
2091	gi1684932	Mycoplasma pneumoniae	adhesin protein	132	35
2091	gi5114063	Mycoplasma pneumoniae	AF090172_1 revertant adhesin-related protein P30	128	35
2092	gi11094019	Homo sapiens	AF305057_2 RTS beta	2047	94
2092	gi1150421	Homo sapiens	rTSbeta	2053	94
2092	gi12654883	Homo sapiens	AAH01285 rTS beta protein	2053	94
2094	gi13432042	Homo sapiens	integrin-linked kinase-associated serine/threonine phosphatase 2C	2018	100
2094	gi16306907	Homo sapiens	AAH06576 integrin-linked	2018	100

257

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			kinase-associated serine/threonine phosphatase 2C		
2094	gi20072498	Mus musculus	0710007A14Rik protein	1935	95
2095	gi18490682	Homo sapiens	fibulin 1	281	37
2095	gi28175169	Mus musculus	1300015B04Rik protein	589	74
2095	gi31419	Homo sapiens	fibulin-1 C	281	37
2096	gi18480746	Mus musculus	olfactory receptor MOR261-10	1336	80
2096	gi21928655	Homo sapiens	seven transmembrane helix receptor	1427	90
2096	gi32052225	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4341246-4340281	1336	80
2097	gi18480746	Mus musculus	olfactory receptor MOR261-10	1336	80
2097	gi21928655	Homo sapiens	seven transmembrane helix receptor	1427	90
2097	gi32052225	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4341246-4340281	1336	80
2098	gi4760780	Mus musculus	Ten-m3	401	95
2098	gi5307761	Danio rerio	ten-m3	347	80
2098	gi6760369	Mus musculus	AF195418_1 ODZ3	401	95
2099	gi21205852	Homo sapiens	AF385429_1 T-cell activation Rho GTPase activating protein; TA-GAP	989	100
2099	gi21410139	Mus musculus	T-cell activation Rho GTPase-activating protein	813	82
2099	gi24980955	Mus musculus	T-cell activation Rho GTPase-activating protein	813	82
2100	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	242	58
2100	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	283	59
2100	gi32486167	Homo sapiens	AD7C-NTP	283	59
2101	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	242	58
2101	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	283	59
2101	gi32486167	Homo sapiens	AD7C-NTP	283	59
2102	gi20196856	Arabidopsis thaliana	putative myosin heavy chain	387	47
2102	gi3142302	Arabidopsis thaliana	Z34293 from A. thaliana.	389	47
2102	gi532124	Dictyostelium discoideum	myosin IC	388	46
2103	gi20196856	Arabidopsis thaliana	putative myosin heavy chain	387	47
2103	gi3142302	Arabidopsis thaliana	Z34293 from A. thaliana.	389	47
2103	gi532124	Dictyostelium discoideum	myosin IC	388	46
2104	gi29564894	Homo sapiens	unnamed protein product	174	39
2104	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	174	39
2104	gi32486167	Homo sapiens	AD7C-NTP	174	39
2105	gi21265163	Homo sapiens		1893	95
2105	gi7248845	Homo sapiens	AF231124_1 testican-1	1893	95
2105	gi793845	Homo sapiens	testican	1893	95
2106	gi12804465	Homo sapiens	AAH01639 prostate cancer	686	66

258

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			overexpressed gene 1		
2106	gi20380774	Homo sapiens		1098	99
2106	gi3462515	Homo sapiens	PB39	686	66
2107	gi12804465	Homo sapiens	AAH01639 prostate cancer overexpressed gene 1	686	66
2107	gi20380774	Homo sapiens		1098	99
2107	gi3462515	Homo sapiens	PB39	686	66
2108	gi17391348	Homo sapiens	AAH18615 Similar to brain expressed, X-linked 1	664	100
2108	gi7689029	Homo sapiens	AF220189_1 uncharacterized hypothalamus protein HBEX2	664	100
2108	gi9963771	Homo sapiens	AF183416_1 ovarian granulosa cell 13.0 kDa protein hGR74 homolog	664	100
2109	gi26353296	Mus musculus	unnamed protein product	711	76
2109	gi28799187	Homo sapiens	unnamed protein product	1463	98
2109	gi30908853	Homo sapiens	synleurin	1463	98
2111	gi20988071	Mus musculus	2600011E07Rik protein	445	89
2111	gi23274133	Homo sapiens	Similar to serine/arginine repetitive matrix 1	161	27
2111	gi3153821	Mus musculus	plenty-of-prolines-101; POP101; SH3-philo-protein	164	30
2112	gi9651079	Macaca fascicularis	hypothetical protein	291	75
2113	gi12408272	Homo sapiens	apolipoprotein L-IV splice variant a	1726	99
2113	gi12408286	Homo sapiens	apolipoprotein L-IV splice variant a	1726	99
2113	gi13374351	Homo sapiens	AF305226_1 apolipoprotein L4	1709	98
2114	gi12408272	Homo sapiens	apolipoprotein L-IV splice variant a	1726	99
2114	gi12408286	Homo sapiens	apolipoprotein L-IV splice variant a	1726	99
2114	gi13374351	Homo sapiens	AF305226_1 apolipoprotein L4	1709	98
2115	gi21744725	Homo sapiens	AF478693_1 glycosyl-phosphatidyl-inositol-MAM	717	97
2115	gi25005318	Sus scrofa	MAM domain containing glycosylphosphatidylinositol anchor 1	672	91
2115	gi25005320	Sus scrofa	glycosylphosphatidylinositol anchor 1 protein	672	91
2116	gi21744725	Homo sapiens	AF478693_1 glycosyl-phosphatidyl-inositol-MAM	717	97
2116	gi25005318	Sus scrofa	MAM domain containing glycosylphosphatidylinositol anchor 1	672	91
2116	gi25005320	Sus scrofa	glycosylphosphatidylinositol anchor 1 protein	672	91
2117	gi16769264	Drosophila melanogaster	LD21615p	219	40
2117	gi7290426	Drosophila melanogaster	CG2875-PB	219	40
2117	gi7290427	Drosophila melanogaster	CG2875-PA	219	40

259

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2118	gi23273399	Homo sapiens		963	100
2118	gi25059032	Mus musculus		686	72
2118	gi28385965	Mus musculus	Similar to phospholipase A2	488	77
2119	gi23273399	Homo sapiens		963	100
2119	gi25059032	Mus musculus		686	72
2119	gi28385965	Mus musculus	Similar to phospholipase A2	488	77
2120	gi13562004	Nephila madagascariensis	AF350276_1 major ampullate spidroin 2-like protein	228	27
2120	gi13562008	Nephila madagascariensis	AF350278_1 major ampullate spidroin 2	238	29
2120	gi159714	Nephila clavipes	dragline silk fibroin	224	29
2121	gi13161409	Mus musculus	family 4 cytochrome P450	445	76
2121	gi13182964	Mus musculus	AF233643_1 cytochrome P450 CYP4F13	191	38
2121	gi13278244	Mus musculus	cytochrome P450, family 4, subfamily f, polypeptide 13	191	38
2122	gi10944887	Homo sapiens	FGFR-like protein	1858	97
2122	gi13183618	Homo sapiens	AF312678_1 FGF homologous factor receptor	1807	96
2122	gi13447749	Homo sapiens	AF279689_1 fibroblast growth factor receptor 5	1858	97
2123	gi10944887	Homo sapiens	FGFR-like protein	1858	97
2123	gi13183618	Homo sapiens	AF312678_1 FGF homologous factor receptor	1807	96
2123	gi13447749	Homo sapiens	AF279689_1 fibroblast growth factor receptor 5	1858	97
2124	gi10944887	Homo sapiens	FGFR-like protein	1858	97
2124	gi13183618	Homo sapiens	AF312678_1 FGF homologous factor receptor	1807	96
2124	gi13447749	Homo sapiens	AF279689_1 fibroblast growth factor receptor 5	1858	97
2125	gi12667454	Rattus norvegicus	AF336858_1 synaptotagmin VIIc	949	88
2125	gi12667456	Rattus norvegicus	AF336859_1 synaptotagmin VIId	949	88
2125	gi12667458	Rattus norvegicus	AF336860_1 synaptotagmin VIIe	949	88
2126	gi12053709	Homo sapiens	with thrombospondin type 1 motif, 12	1143	98
2126	gi27817773	Mus musculus	metalloprotease disintegrin 12 protein	873	76
2126	gi5923788	Homo sapiens	AF140675_1 zinc metalloprotease ADAMTS7	271	39
2127	gi11493982	Homo sapiens	AF208232_1 TLH29 protein precursor	303	70
2127	gi15929988	Homo sapiens	AAH15423 Similar to TLH29 protein precursor	497	100
2127	gi21618549	Homo sapiens	TLH29 protein precursor	303	70
2128	gi17391206	Mus musculus	RIKEN cDNA 2210412D01	1267	99
2128	gi23468210	Homo sapiens	Similar to CGI-67 protein	1096	81
2128	gi9368522	Homo sapiens	CGI-67 protein	1267	99
2129	gi17391206	Mus musculus	RIKEN cDNA 2210412D01	1267	99
2129	gi23468210	Homo sapiens	Similar to CGI-67 protein	1096	81
2129	gi9368522	Homo sapiens	CGI-67 protein	1267	99

260

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2130	gi20071312	Mus musculus	4933425F03Rik protein	614	85
2130	gi33391740	Homo sapiens	MGC45780	426	96
2130	gi735	Bos taurus	scavenger receptor type I	336	51
2131	gi20071312	Mus musculus	4933425F03Rik protein	614	85
2131	gi33391740	Homo sapiens	MGC45780	426	96
2131	gi735	Bos taurus	scavenger receptor type I	336	51
2132	gi5870866	Homo sapiens	TATA element modulatory factor	4531	99
2132	gi6650548	Rattus norvegicus	AF107843_1 TATA element modulatory factor	2583	82
2132	gi7290766	Drosophila melanogaster	CG4557-PA	692	25
2133	gi1020145	Homo sapiens	DNA binding protein	1483	43
2133	gi18643896	Homo sapiens	zinc finger protein	1486	43
2133	gi29476835	Homo sapiens		1486	43
2134	gi16198520	Homo sapiens	Saccharomyces cerevisiae Nip7p homolog	944	100
2134	gi4680713	Homo sapiens	AF132971_1 CGI-37 protein	944	100
2134	gi5114055	Homo sapiens	HSPC031	944	100
2135	gi23274241	Homo sapiens	KIAA1892-like	563	86
2135	gi26332114	Mus musculus	unnamed protein product	577	89
2135	gi26345386	Mus musculus	unnamed protein product	577	89
2136	gi15620885	Homo sapiens	KIAA1913 protein	1627	99
2136	gi26339494	Mus musculus	unnamed protein product	1480	90
2136	gi28279830	Homo sapiens	KIAA1913 protein	1598	99
2137	gi1000448	Rattus norvegicus	Rat kidney AGT2 precursor	1578	84
2137	gi12406973	Homo sapiens	alanine-glyoxylate aminotransferase 2	1865	98
2137	gi1944136	Rattus norvegicus	beta-alanine-pyruvate aminotransferase	1625	85
2138	gi1000448	Rattus norvegicus	Rat kidney AGT2 precursor	1578	84
2138	gi12406973	Homo sapiens	alanine-glyoxylate aminotransferase 2	1865	98
2138	gi1944136	Rattus norvegicus	beta-alanine-pyruvate aminotransferase	1625	85
2139	gi29436673	Mus musculus	1700049K14Rik protein	648	100
2139	gi4204421	Euroglyphus maynei	group 3 allergen Eur m 3 0101 precursor	212	40
2139	gi5441861	Paralichthys olivaceus	chymotrypsinogen 2	210	36
2140	gi17985046	Brucella melitensis 16M	GLYCOSYL TRANSFERASE	130	28
2140	gi20515259	Thermoanaerobacter tengcongensis	predicted glycosyltransferases	133	32
2140	gi4455730	Streptomyces coelicolor A3(2)	putative transferase	140	32
2141	gi13649477	Homo sapiens	AF250309_1 putative cytokine receptor CRL4 precursor	2694	100
2141	gi30584223	synthetic construct	Homo sapiens interleukin 17B receptor	2694	100
2141	gi9246433	Homo sapiens	AF208110_1 IL-17 receptor homolog precursor	2688	99
2142	gi18676472	Homo sapiens	FLJ00133 protein	855	76
2142	gi29568116	Mus musculus	secreted protein SST3	725	64

261

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2142	gi499686	Heliocidaris erythrogramma	fibropellin Ia	390	40
2143	gi16588687	Homo sapiens	AF315687_1 S-adenosylhomocysteine hydrolase-like protein	147	100
2143	gi27692283	Mus musculus	S-adenosylhomocysteine hydrolase-like 1	147	100
2143	gi2852125	Homo sapiens	S-adenosyl homocysteine hydrolase homolog	147	100
2144	gi16740861	Homo sapiens	AAH16292 ubiquitin-conjugating enzyme E2C	521	66
2144	gi29791813	Homo sapiens	Ubiquitin-conjugating enzyme E2C, isoform 1	521	66
2144	gi30583439	Homo sapiens	ubiquitin-conjugating enzyme E2C	521	66
2145	gi20086516	Homo sapiens	AF245303_1 prominin-2 variant A	2480	91
2145	gi20086518	Homo sapiens	AF245304_1 prominin-2 variant B	2480	91
2145	gi24637566	Rattus norvegicus	prominin-2	1876	68
2146	gi29351676	Homo sapiens	Angiopoietin-like 5	1310	99
2146	gi29468510	Homo sapiens	putative fibrinogen-like protein	1305	99
2146	gi9229906	Ciona intestinalis	fibrinogen-like protein	392	39
2147	gi29351676	Homo sapiens	Angiopoietin-like 5	1310	99
2147	gi29468510	Homo sapiens	putative fibrinogen-like protein	1305	99
2147	gi9229906	Ciona intestinalis	fibrinogen-like protein	392	39
2148	gi29351676	Homo sapiens	Angiopoietin-like 5	1310	99
2148	gi29468510	Homo sapiens	putative fibrinogen-like protein	1305	99
2148	gi9229906	Ciona intestinalis	fibrinogen-like protein	392	39
2150	gi13543706	Homo sapiens	AAH06003	349	100
2150	gi20988061	Mus musculus	1810013D10Rik protein	333	92
2150	gi21619079	Homo sapiens		349	100
2151	gi11493652	Homo sapiens	AF200708_1 calcium channel blocker resistance protein CCB1	2168	100
2151	gi13924720	Homo sapiens	AF252872_1 cystine/glutamate transporter xCT	2168	100
2151	gi15082352	Homo sapiens	AAH12087 member 11	2168	100
2152	gi18043214	Mus musculus	serine/arginine-rich protein specific kinase 2	132	67
2152	gi23270876	Homo sapiens	Similar to SFRS protein kinase 2	132	67
2152	gi3406050	Homo sapiens	serine kinase SRPK2	132	67
2153	gi22164066	Homo sapiens	AF388385_1 neuroblastoma-amplified protein	4284	99
2153	gi30353863	Homo sapiens	NAG protein	4298	99
2153	gi4337460	Homo sapiens	neuroblastoma-amplified protein	4272	99
2154	gi22164066	Homo sapiens	AF388385_1 neuroblastoma-amplified protein	4284	99
2154	gi30353863	Homo sapiens	NAG protein	4298	99
2154	gi4337460	Homo sapiens	neuroblastoma-amplified protein	4272	99
2155	gi1008367	Saccharomyces cerevisiae	CPS1	131	48

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2155	gi3594	Saccharomyces cerevisiae	carboxypeptidase s	131	48
2155	gi3596	Saccharomyces cerevisiae	carboxypeptidase yscS	131	48
2156	gi11558029	Homo sapiens	organic cation transporter	1876	100
2156	gi18088251	Homo sapiens	AAH20565 Similar to hBOIT for potent brain type organic ion transporter	1838	95
2156	gi9663117	Homo sapiens	organic cation transporter	1868	99
2157	gi21732438	Homo sapiens	hypothetical protein	567	100
2157	gi26330392	Mus musculus	unnamed protein product	486	85
2157	gi26390211	Mus musculus	unnamed protein product	486	85
2158	gi23893591	Human herpesvirus 4	BHFLF1 early reading frame	169	28
2158	gi30844300	Cercopithecine herpesvirus 1	immediate early protein ICP0	166	23
2158	gi30844317	Cercopithecine herpesvirus 1	immediate early protein ICP0	166	23
2159	gi27804346	Homo sapiens	BRD4-NUT fusion oncoprotein	3773	99
2159	gi3115204	Homo sapiens	HUNKI	3787	99
2159	gi3184498	Homo sapiens	R31546_1	3837	99
2160	gi15420832	Homo sapiens	AF397394_1 NOE3-3	535	96
2160	gi15420834	Homo sapiens	AF397395_1 NOE3-4	535	96
2160	gi18490927	Homo sapiens	olfactomedin 3	531	95
2161	gi22209078	Homo sapiens	hypothetical protein DKFZp566D234	773	98
2161	gi6330966	Homo sapiens	KIAA1263 protein	773	98
2161	gi6808053	Homo sapiens	hypothetical protein	766	97
2162	gi12654031	Homo sapiens	AAH00819 Similar to CG6950 gene product	158	93
2162	gi21707106	Homo sapiens		120	56
2162	gi758591	Homo sapiens	glutamine--phenylpyruvate aminotransferase	120	56
2163	gi21666433	Mus musculus	AF404775_1 actin-binding LIM protein 1 medium isoform	302	54
2163	gi2337952	Homo sapiens	actin-binding double-zinc-finger protein	303	54
2163	gi30259308	Mus musculus	actin-binding LIM protein 2	498	79
2164	gi2062399	Rattus norvegicus	protein serine/threonine kinase CPG16	404	50
2164	gi6716518	Mus musculus	AF1551 doublecortin-like kinase	404	50
2164	gi6716522	Mus musculus	AF155821_1 CPG16	404	50
2165	gi2062399	Rattus norvegicus	protein serine/threonine kinase CPG16	404	50
2165	gi6716518	Mus musculus	AF1551 doublecortin-like kinase	404	50
2165	gi6716522	Mus musculus	AF155821_1 CPG16	404	50
2166	gi13436035	Mus musculus	prostaglandin E synthase 2	1321	87
2166	gi29179467	Danio rerio	Similar to prostaglandin E synthase 2	988	66
2166	gi9280108	Macaca fascicularis	membrane-associated prostaglandin E synthase-2	1449	97
2167	gi12805247	Mus musculus	Complement component 1, q subcomponent, alpha	955	70

263

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			polypeptide		
2167	gi20988805	Homo sapiens	complement component 1, q subcomponent, alpha polypeptide	1318	100
2167	gi4894854	Homo sapiens	AF135157_1 complement C1q A chain precursor	1318	100
2168	gi1491621	Bovine herpesvirus 1	UL36	126	38
2168	gi15145795	Sus scrofa	basic proline-rich protein	123	38
2168	gi2653311	Bovine herpesvirus type 1.1 (strain Cooper)		126	38
2169	gi21707458	Homo sapiens	PAX transcription activation domain interacting protein 1 like	2470	81
2169	gi2565046	Homo sapiens	CAGF28	3770	97
2169	gi4336734	Mus musculus	Pax transcription activation domain interacting protein PTIP	2945	70
2170	gi21707458	Homo sapiens	PAX transcription activation domain interacting protein 1 like	2470	81
2170	gi2565046	Homo sapiens	CAGF28	3770	97
2170	gi4336734	Mus musculus	Pax transcription activation domain interacting protein PTIP	2945	70
2171	gi32488718	Oryza sativa (japonica cultivar-group)	OSJNBa0088H09.19	121	41
2172	gi26353296	Mus musculus	unnamed protein product	711	76
2172	gi28799187	Homo sapiens	unnamed protein product	1463	98
2172	gi30908853	Homo sapiens	synleurin	1463	98
2173	gi13991167	Homo sapiens	sialic acid-binding immunoglobulin-like lectin-like long splice variant	1231	99
2173	gi14625822	Homo sapiens	AF282256_1 Siglec-L1	1231	99
2173	gi23272769	Homo sapiens	SIGLEC-like 1	1231	99
2174	gi13435476	Mus musculus	DNA segment, Chr 10, University of California at Los Angeles 1	1206	91
2174	gi28279553	Danio rerio	Similar to DNA segment, Chr 10, University of California at Los Angeles 1	865	69
2174	gi29144983	Mus musculus	DNA segment, Chr 6, ERATO Doi 253, expressed	668	67
2175	gi27924102	Mus musculus	2310075M15Rik protein	944	68
2175	gi29436830	Mus musculus	2310075M15Rik protein	944	68
2175	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	940	67
2176	gi27924102	Mus musculus	2310075M15Rik protein	944	68
2176	gi29436830	Mus musculus	2310075M15Rik protein	944	68
2176	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	940	67
2177	gi27924102	Mus musculus	2310075M15Rik protein	944	68
2177	gi29436830	Mus musculus	2310075M15Rik protein	944	68
2177	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	940	67
2178	gi11493483	Homo sapiens	AF130117_48 PRO2550	220	56

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2178	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	220	51
2178	gi8572229	Homo sapiens	ubiquitous TPR-motif protein Y isoform	217	53
2179	gi6808611	Homo sapiens	AF204231_1 88-kDa Golgi protein	3209	97
2179	gi6969980	Homo sapiens	AF163441_1 golgin 67	2339	98
2179	gi7211438	Homo sapiens	AF164622_1 golgin-67	2321	97
2180	gi15030299	Mus musculus	protein kinase, cAMP dependent regulatory, type I beta	1881	94
2180	gi200365	Mus musculus	cAMP-dependent protein kinase regulatory subunit	1886	94
2180	gi307377	Homo sapiens	cAMP-dependent protein kinase RI-beta regulatory subunit	1957	99
2181	gi10945428	Homo sapiens	membrane-associated guanylate kinase MAGI3	156	41
2181	gi12003994	Homo sapiens	AF213259_1 membrane-associated guanylate kinase-related MAGI-3	156	41
2181	gi7650497	Rattus norvegicus	AF255614_1 scaffolding protein SLIPR	156	41
2182	gi1845577	Mus musculus	-lipoxygenase	2559	74
2182	gi30047223	Mus musculus	Arachidonate lipoxygenase, epidermal	2557	74
2182	gi3645913	Mus musculus	-lipoxygenase	2559	74
2183	gi1845577	Mus musculus	-lipoxygenase	2559	74
2183	gi30047223	Mus musculus	Arachidonate lipoxygenase, epidermal	2557	74
2183	gi3645913	Mus musculus	-lipoxygenase	2559	74
2184	gi1845577	Mus musculus	-lipoxygenase	2559	74
2184	gi30047223	Mus musculus	Arachidonate lipoxygenase, epidermal	2557	74
2184	gi3645913	Mus musculus	-lipoxygenase	2559	74
2185	gi10439485	Homo sapiens	unnamed protein product	481	87
2185	gi12853469	Mus musculus	unnamed protein product	395	62
2185	gi18027736	Homo sapiens	AF318322_1 unknown	330	50
2186	gi14198207	Mus musculus	hypothetical protein BC008163	1599	98
2186	gi19343692	Homo sapiens		1625	100
2186	gi7294965	Drosophila melanogaster	CG4452-PA	615	40
2192	gi22209089	Homo sapiens	Similar to vesicular inhibitory amino acid transporter	308	98
2192	gi30354125	Mus musculus	Viaat protein	308	98
2192	gi31566392	Homo sapiens	Vesicular inhibitory amino acid transporter	308	98
2193	gi22507470	Mus musculus	AI413481 protein	997	92
2193	gi3097285	Rattus norvegicus	ZOG	481	48
2193	gi802014	Rattus norvegicus	preadipocyte factor 1	481	48
2194	gi1488314	Homo sapiens	hepatitis delta antigen interacting protein A	442	49
2194	gi18088059	Mus musculus	E030025D05Rik protein	1622	83
2194	gi6624073	Homo sapiens	AC007743_1 similar to hepatitis delta antigen	1903	94

265

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			interacting protein A		
2195	gi14250638	Homo sapiens	AAH08783 Similar to DNA segment, Chr 17, human D6S54E	1886	99
2195	gi3941733	Mus musculus	AAC82476 BAT4	1453	76
2195	gi4337106	Homo sapiens	AAD18082 BAT4	1886	99
2196	gi15277895	Homo sapiens	AAH12939 Similar to cardiotrophin-like cytokine; neurotrophin-1/B-cell stimulating factor-3	1226	100
2196	gi16356643	Homo sapiens	cardiotrophin-like cytokine	1226	100
2196	gi6007643	Homo sapiens	neurotrophin-1/B-cell stimulating factor-3	1226	100
2197	gi15982236	Mus musculus	putative methionyl aminopeptidase	1069	92
2197	gi23306398	Arabidopsis thaliana	, putative	739	50
2197	gi24899771	Arabidopsis thaliana	, putative	739	50
2198	gi13592175	Leishmania major	AC084329_1 ppg3	196	24
2198	gi28828184	Dictyostelium discoideum	similar to Leishmania major. Ppg3	180	24
2198	gi5420387	Leishmania major	proteophosphoglycan	202	24
2199	gi19387136	Homo sapiens	AF479748_1 PYRIN-containing APAF1-like protein 5	4151	91
2199	gi21410402	Mus musculus	PYRIN-containing APAF1-like protein 5	1191	54
2199	gi28436366	Homo sapiens	NALP6	4151	91
2200	gi11321325	Homo sapiens	AF311862_1 Lin-7b	684	98
2200	gi20381193	Homo sapiens	Lin-7b protein; likely ortholog of mouse LIN-7B; mammalian LIN-7 protein 2	684	98
2200	gi3885828	Rattus norvegicus	lin-7-A	673	96
2201	gi14349125	Homo sapiens	alpha2-glucosyltransferase	567	97
2201	gi32490259	Oryza sativa (japonica cultivar-group)	OSJNBb0116K07.1	181	46
2201	gi3513451	Rattus norvegicus	potassium channel regulator 1	549	96
2202	gi13325140	Homo sapiens	AAH04383	2693	100
2202	gi35768	Homo sapiens	polypirimidine tract binding protein	2693	100
2202	gi35774	Homo sapiens		2693	100
2203	gi21522776	Homo sapiens	unnamed protein product	2998	98
2203	gi24047224	Homo sapiens	Similar to EGF-like-domain, multiple 6	2982	98
2203	gi6752658	Homo sapiens	AF186084_1 epidermal growth factor repeat containing protein	2984	98
2204	gi21522776	Homo sapiens	unnamed protein product	2998	98
2204	gi24047224	Homo sapiens	Similar to EGF-like-domain, multiple 6	2982	98
2204	gi6752658	Homo sapiens	AF186084_1 epidermal growth factor repeat containing protein	2984	98
2205	gi11385648	Homo sapiens	AF273045_1 CTCL tumor antigen se14-3	3622	95
2205	gi17980969	Homo sapiens	AF454056_1 se14-3r protein	3858	95
2205	gi29165763	Mus musculus	3632413B07Rik protein	3261	75

266

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2206	gi11385648	Homo sapiens	AF273045_1 CTCL tumor antigen se14-3	3622	95
2206	gi17980969	Homo sapiens	AF454056_1 se14-3r protein	3858	95
2206	gi29165763	Mus musculus	3632413B07Rik protein	3261	75
2207	gi11385648	Homo sapiens	AF273045_1 CTCL tumor antigen se14-3	3622	95
2207	gi17980969	Homo sapiens	AF454056_1 se14-3r protein	3858	95
2207	gi29165763	Mus musculus	3632413B07Rik protein	3261	75
2208	gi11385648	Homo sapiens	AF273045_1 CTCL tumor antigen se14-3	3622	95
2208	gi17980969	Homo sapiens	AF454056_1 se14-3r protein	3858	95
2208	gi29165763	Mus musculus	3632413B07Rik protein	3261	75
2209	gi14043211	Homo sapiens	AAH07594 Similar to RIKEN cDNA 4931428F04 gene	975	97
2209	gi21750866	Homo sapiens	unnamed protein product	975	97
2209	gi25058997	Mus musculus	1110003N12Rik protein	641	62
2210	gi19387136	Homo sapiens	AF479748_1 PYRIN-containing APAF1-like protein 5	3078	100
2210	gi202806	Rattus norvegicus	vasopressin receptor	969	67
2210	gi28436366	Homo sapiens	NALP6	3078	100
2211	gi13157560	Homo sapiens		2246	99
2211	gi18147612	Homo sapiens	metalloprotease disintegrin	2246	99
2211	gi21908030	Homo sapiens	a disintegrin and metalloprotease domain 33	2230	98
2212	gi13592175	Leishmania major	AC084329_1 ppg3	163	34
2212	gi15145803	Chlamydomonas reinhardtii	hydroxyproline-rich glycoprotein VSP4	150	28
2212	gi5420387	Leishmania major	proteophosphoglycan	157	32
2213	gi15420879	Mus musculus	AF398971_1 ankyrin repeat-containing SOCS box protein 10	1986	83
2213	gi18031949	Mus musculus	SOCS box protein ASB-18	808	44
2213	gi18092200	Homo sapiens	AF417920_1 ASB-10	2062	91
2214	gi32707	Homo sapiens	interferon-omega 1	331	51
2214	gi386800	Homo sapiens	interferon-alpha	334	51
2214	gi491284	synthetic construct	IFN-pseudo-omega 2	806	99
2215	gi6841550	Homo sapiens	AF161513_1 HSPC164	1594	99
2215	gi6841560	Homo sapiens	AF161518_1 HSPC169	1604	100
2215	gi9844577	Homo sapiens		1601	99
2216	gi11493483	Homo sapiens	AF130117_48 PRO2550	408	79
2216	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	352	74
2216	gi7020440	Homo sapiens	unnamed protein product	396	76
2217	gi22658418	Mus musculus	cDNA sequence BC030934	365	71
2217	gi28838433	Homo sapiens	DKFZp762A2013 protein	443	87
2217	gi30842594	Homo sapiens	putative sulfhydryl oxidase precursor	360	74
2218	gi12958660	Homo sapiens	AF321918_1 acid phosphatase	573	89
2218	gi12958663	Homo sapiens	AF321918_4 acid phosphatase variant 3	573	89
2218	gi202934	Rattus norvegicus		207	43
2219	gi15866260	Homo sapiens	AF411132_1 MRIP2	2479	97
2219	gi29476839	Homo sapiens	Similar to centaurin, gamma 2	2124	98

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2219	gi30354556	Homo sapiens	MRIP2 protein	2466	97
2220	gi15866260	Homo sapiens	AF411132_1 MRIP2	2479	97
2220	gi29476839	Homo sapiens	Similar to centaurin, gamma 2	2124	98
2220	gi30354556	Homo sapiens	MRIP2 protein	2466	97
2221	gi15866260	Homo sapiens	AF411132_1 MRIP2	2479	97
2221	gi29476839	Homo sapiens	Similar to centaurin, gamma 2	2124	98
2221	gi30354556	Homo sapiens	MRIP2 protein	2466	97
2222	gi15866260	Homo sapiens	AF411132_1 MRIP2	2479	97
2222	gi29476839	Homo sapiens	Similar to centaurin, gamma 2	2124	98
2222	gi30354556	Homo sapiens	MRIP2 protein	2466	97
2223	gi1841702	Macaca fascicularis	fertilin alpha-I isoform	655	83
2223	gi2632092	Pongo pygmaeus	fertilin alpha protein	745	94
2223	gi2655944	Papio anubis	fertilin alpha-I	661	85
2224	gi17887359	Oryctolagus cuniculus	lipophilin AL2	248	54
2224	gi4107229	Homo sapiens	lipophilin A	454	100
2224	gi4107231	Homo sapiens	lipophilin B	267	60
2225	gi180251	Homo sapiens	precerebellin	183	48
2225	gi6942096	Mus musculus	CBLN3	472	90
2225	gi6942098	Mus musculus	AF218380_1 CBLN3	472	90
2226	gi18255724	Mus musculus	LOC215928 protein	131	28
2226	gi21750370	Homo sapiens	unnamed protein product	917	85
2226	gi28460663	Rattus norvegicus	Na ⁺ dependent glucose transporter 1	185	30
2227	gi18255724	Mus musculus	LOC215928 protein	131	28
2227	gi21750370	Homo sapiens	unnamed protein product	917	85
2227	gi28460663	Rattus norvegicus	Na ⁺ dependent glucose transporter 1	185	30
2228	gi5726236	multiple sclerosis associated retrovirus element	gag polyprotein	173	53
2228	gi5726238	multiple sclerosis associated retrovirus element	AF123881_1 gag polyprotein	163	57
2228	gi8272464	Homo sapiens	AF156961_1 gag	191	56
2229	gi12964746	Mus musculus	AF316612_1 neuronal pentraxin receptor	2225	88
2229	gi2253263	Rattus norvegicus	neuronal pentraxin receptor	2250	88
2229	gi4160197	Homo sapiens		2559	99
2230	gi3170615	Mus musculus	DOC4	1520	95
2230	gi4760782	Mus musculus	Ten-m4	1520	95
2230	gi9909617	Gallus gallus	teneurin-4	1333	89
2232	gi14124993	Homo sapiens		232	83
2232	gi30704639	Mus musculus	4930553F24Rik protein	210	74
2232	gi7716100	Rattus norvegicus	AF226993_1 selective LIM binding factor	213	76
2233	gi20987535	Mus musculus	Mcoln2 protein	804	92
2233	gi24417793	Mus musculus	mucolipin 2	804	92
2233	gi24417795	Homo sapiens	mucolipin 2	857	99
2234	gi20987535	Mus musculus	Mcoln2 protein	804	92
2234	gi24417793	Mus musculus	mucolipin 2	804	92
2234	gi24417795	Homo sapiens	mucolipin 2	857	99
2235	gi22477432	Homo sapiens	DKFZP762N2316 protein	1002	100
2235	gi27370669	Homo sapiens	Similar to RE1-silencing	159	36

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			transcription factor		
2235	gi403020	Mus musculus	En-2/lacZ fusion protein	330	92
2238	gi11990126	Camelus dromedarius	chymosin	294	83
2238	gi491952	synthetic construct	preprochymosin	291	83
2238	gi7008025	Callithrix jacchus	prochymosin	314	91
2239	gi27356934	Homo sapiens	extracellular sulfatase SULF-2	560	100
2239	gi27356938	Mus musculus	extracellular sulfatase SULF-2	499	90
2239	gi29165845	Mus musculus	Extracellular sulfatase SULF-1	375	70
2240	gi27124671	Homo sapiens	Zn-carboxypeptidase	877	96
2240	gi2960072	Homo sapiens	procarboxypeptidase B	488	55
2240	gi32880163	Homo sapiens		487	55
2241	gi27124671	Homo sapiens	Zn-carboxypeptidase	877	96
2241	gi2960072	Homo sapiens	procarboxypeptidase B	488	55
2241	gi32880163	Homo sapiens		487	55
2242	gi11545705	Homo sapiens	ISCU1	663	99
2242	gi11545707	Homo sapiens	ISCU2	845	100
2242	gi20381021	Mus musculus	Nifu-pending protein	807	96
2243	gi17512406	Mus musculus	differential display and activated by p53	188	52
2243	gi25166615	Homo sapiens	AF223000_1 DDA3-like protein	427	56
2243	gi25166621	Homo sapiens	AF322891_1 DDA3-like protein	427	56
2244	gi15990480	Homo sapiens	-binding protein 2	1200	99
2244	gi21961217	Homo sapiens	-binding protein 2	1200	99
2244	gi22213050	Mus musculus	B230313N05Rik protein	1189	97
2245	gi204058	Rattus norvegicus	extracellular signal-related kinase 3	1497	62
2245	gi23903	Homo sapiens	63kDa protein kinase	2886	98
2245	gi27882123	Danio rerio	Similar to mitogen-activated protein kinase 4	1670	61
2246	gi24417711	Homo sapiens	nesprin-2	354	100
2246	gi28195679	Homo sapiens	nesprin-2 alpha 2	354	100
2246	gi28195681	Homo sapiens	nesprin-2 beta 2	354	100
2248	gi19353133	Mus musculus	C1q-like	560	80
2248	gi26996600	Mus musculus	Similar to C1q-like	692	96
2248	gi32401227	Homo sapiens	AF525315_1 C1q-domain containing protein	711	99
2249	gi14718648	Homo sapiens	allantoicase	967	99
2249	gi20987689	Homo sapiens	Similar to allantoicase	1162	99
2249	gi9255889	Mus musculus	AF278712_1 allantoicase	932	78
2250	gi15617341	Homo sapiens	LAG-3 protein precursor	2796	99
2250	gi30851187	Homo sapiens	LAG3 protein	1906	99
2250	gi579596	Homo sapiens	lymphocyte protein	2634	98
2251	gi13810285	Rattus norvegicus	guanine nucleotide release/exchange factor	5807	91
2251	gi2522208	Homo sapiens	Ras-GRF2	6407	99
2251	gi5882290	Homo sapiens	Ras guanine nucleotide exchange factor 2	6401	99
2252	gi22038159	Homo sapiens	AF527605_1 zizimin1	7984	100
2252	gi28374168	Mus musculus	AA959601 protein	7520	93
2252	gi31419757	Mus musculus	AA959601 protein	7520	93
2253	gi10433672	Homo sapiens	unnamed protein product	1325	89

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2253	gi19263505	Homo sapiens	hypothetical protein FLJ12242	1325	89
2253	gi23272394	Homo sapiens	KCTD2 protein	728	67
2254	gi14041697	Homo sapiens		3330	94
2254	gi21594273	Homo sapiens		3371	95
2254	gi25303955	Homo sapiens		3371	95
2255	gi1438532	Rattus norvegicus	rA1	393	51
2255	gi1438534	Rattus norvegicus	rA9	857	70
2255	gi9438033	Homo sapiens	AF254411_1 ser/arg-rich pre-mRNA splicing factor SR-A1	386	51
2256	gi1438532	Rattus norvegicus	rA1	393	51
2256	gi1438534	Rattus norvegicus	rA9	857	70
2256	gi9438033	Homo sapiens	AF254411_1 ser/arg-rich pre-mRNA splicing factor SR-A1	386	51
2257	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	242	58
2257	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	283	59
2257	gi32486167	Homo sapiens	AD7C-NTP	283	59
2258	gi12652851	Homo sapiens	AAH00178 potassium channel modulatory factor	1987	100
2258	gi26453336	Homo sapiens	FIGC1	1983	99
2258	gi7677058	Homo sapiens	AF155652_1 potassium channel modulatory factor	1983	99
2259	gi27695389	Mus musculus	MGC58017 protein	1050	97
2259	gi28558964	Human herpesvirus 4 type 2	nuclear antigen-3B	138	28
2259	gi30481648	Homo sapiens		660	55
2260	gi11119239	Rattus norvegicus	AF313453_1 synaptotagmin 13	792	86
2260	gi14210274	Rattus norvegicus	AF375466_1 synaptotagmin 13	792	86
2260	gi21410154	Mus musculus	synaptotagmin 13	779	84
2261	gi11342591	Mus musculus	RanBP7/importin 7	5301	97
2261	gi32330683	Mus musculus	importin 7	5313	97
2261	gi3800881	Homo sapiens	RanBP7/importin 7	5333	98
2262	gi17939650	Homo sapiens	AAH19302 hypothetical protein FLJ12525	3660	97
2262	gi18676522	Homo sapiens	FLJ00158 protein	1599	100
2262	gi27462078	Homo sapiens	AF116730_1 MSTP060	3629	94
2263	gi28981429	Mus musculus	Ddef1 protein	879	94
2263	gi4063614	Mus musculus	ADP-ribosylation factor-directed GTPase activating protein isoform a	879	94
2263	gi4406393	Bos taurus	differentiation enhancing factor 1	876	94
2264	gi59500	Human herpesvirus 1	RL2	139	37
2264	gi59557	Human herpesvirus 1	immediate early protein	139	37
2264	gi59833	Human herpesvirus 1	IE110	139	37
2265	gi13872813	Homo sapiens	fibulin-6	513	29
2265	gi14575679	Homo sapiens	AF156100_1 hemicentin	513	29
2265	gi9280405	Homo sapiens	AF245505_1 adlcan	1462	46
2266	gi15145797	Sus scrofa	basic proline-rich protein	178	25
2266	gi27348769	Bradyrhizobium japonicum USDA 110	blr0521	191	29
2266	gi30844278	Cercopithecine	very large tegument protein	178	25

270

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
		herpesvirus 1			
2267	gi21748983	Homo sapiens	unnamed protein product	128	65
2267	gi522145	Homo sapiens	B-cell growth factor	129	71
2268	gi21748983	Homo sapiens	unnamed protein product	128	65
2268	gi522145	Homo sapiens	B-cell growth factor	129	71
2269	gi13529248	Homo sapiens	Centrin 3	842	100
2269	gi30582215	Homo sapiens		842	100
2269	gi30584861	synthetic construct		842	100
2270	gi31455256	Homo sapiens	IMAGE3510317 protein	2259	91
2270	gi32492907	Homo sapiens	selenoprotein O	2259	91
2270	gi6572230	Homo sapiens		1768	98
2271	gi31455256	Homo sapiens	IMAGE3510317 protein	2259	91
2271	gi32492907	Homo sapiens	selenoprotein O	2259	91
2271	gi6572230	Homo sapiens		1768	98
2272	gi21928729	Homo sapiens	seven transmembrane helix receptor	661	99
2272	gi6693701	Homo sapiens	AF147788_1 melanopsin	661	99
2272	gi6693703	Mus musculus	AF147789_1 melanopsin	529	83
2273	gi20072741	Mus musculus	E430025L02Rik protein	538	81
2273	gi2104856	Rattus norvegicus	platelet glycoprotein V	143	41
2273	gi439296	Homo sapiens	garp	166	43
2274	gi15487302	Homo sapiens	medium-chain acyl-CoA synthetase	727	97
2274	gi15706421	Homo sapiens	middle-chain acyl-CoA synthetase1	727	97
2274	gi5019275	Bos taurus	xenobiotic/medium-chain fatty acid:CoA ligase form XL-III	529	70
2275	gi15077826	Homo sapiens	AF394782_1 rap guanine nucleotide exchange factor	2149	100
2275	gi20386206	Homo sapiens	AF478567_1 PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2	2149	100
2275	gi6650766	Homo sapiens	AF117947_1 PDZ domain-containing guanine nucleotide exchange factor I	2149	100
2276	gi15077826	Homo sapiens	AF394782_1 rap guanine nucleotide exchange factor	2149	100
2276	gi20386206	Homo sapiens	AF478567_1 PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2	2149	100
2276	gi6650766	Homo sapiens	AF117947_1 PDZ domain-containing guanine nucleotide exchange factor I	2149	100
2277	gi13592175	Leishmania major	AC084329_1 ppg3	165	29
2277	gi5420387	Leishmania major	proteophosphoglycan	163	26
2277	gi5420389	Leishmania major	proteophosphoglycan	151	30
2278	gi18676788	Homo sapiens	unnamed protein product	875	88
2278	gi21779866	Mus musculus	AF458068_1 IL-17RE	234	38
2278	gi21779869	Homo sapiens	AF458069_1 IL-17RE	875	88
2279	gi18676788	Homo sapiens	unnamed protein product	875	88
2279	gi21779866	Mus musculus	AF458068_1 IL-17RE	234	38
2279	gi21779869	Homo sapiens	AF458069_1 IL-17RE	875	88
2280	gi14150450	Rattus norvegicus	AF241241_1 UDP-GalNAc:polypeptide N-	197	85

271

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			acetylgalactosaminyltransferase T9		
2280	gi25809274	Homo sapiens	polypeptide N-acetylgalactosaminyltransferase 10	219	97
2280	gi28268676	Homo sapiens	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10	219	97
2281	gi17384577	Escherichia coli	orf1176	1087	99
2281	gi28629348	Escherichia coli	SopA	1087	99
2281	gi42431	Escherichia coli		1087	99
2282	gi1377895	Homo sapiens	OB-cadherin-2	540	51
2282	gi30171995	Homo sapiens	cadherin-24	990	100
2282	gi30171998	Homo sapiens	cadherin-24 variant	990	100
2283	gi1377895	Homo sapiens	OB-cadherin-2	540	51
2283	gi30171995	Homo sapiens	cadherin-24	990	100
2283	gi30171998	Homo sapiens	cadherin-24 variant	990	100
2284	gi1398903	Mus musculus	Ca ²⁺ dependent activator protein for secretion	1303	89
2284	gi21541504	Homo sapiens	AF458662_1 calcium-dependent activator protein for secretion protein	1185	83
2284	gi577428	Rattus norvegicus	Ca ²⁺ -dependent activator protein; calcium-dependent actin-binding protein	1247	85
2285	gi11071729	Homo sapiens	putative dipeptidase	526	100
2285	gi11125344	Homo sapiens	putative metallopeptidase	263	58
2285	gi32490515	Mus musculus	putative membrane-bound dipeptidase-3	245	55
2286	gi11493652	Homo sapiens	AF200708_1 calcium channel blocker resistance protein CCB1	2168	100
2286	gi13924720	Homo sapiens	AF252872_1 cystine/glutamate transporter xCT	2168	100
2286	gi15082352	Homo sapiens	AAH12087 member 11	2168	100
2287	gi17028348	Homo sapiens	DKFZP586G1517 protein	3748	100
2287	gi20987924	Mus musculus	2410004L15Rik protein	3473	92
2287	gi29612455	Mus musculus	2410004L15Rik protein	3819	92
2288	gi19352987	Homo sapiens	Similar to KIAA0433 protein	6283	97
2288	gi2887437	Homo sapiens	KIAA0433	6416	98
2288	gi31418648	Mus musculus		4916	95
2289	gi24061707	Mus musculus	GAP-related interacting partner to E12	766	88
2289	gi26334941	Mus musculus	unnamed protein product	783	89
2289	gi4240257	Homo sapiens	KIAA0884 protein	725	75
2290	gi20269957	Sus scrofa	AF498759_1 phospholipase C delta 4	166	96
2290	gi21307610	Mus musculus	phospholipase C delta 4	158	90
2290	gi571466	Rattus norvegicus	phospholipase C delta-4	151	84
2291	gi12839717	Mus musculus	unnamed protein product	238	62
2291	gi16552885	Homo sapiens	unnamed protein product	382	92
2291	gi26327387	Mus musculus	unnamed protein product	238	62
2292	gi18480186	Mus musculus	olfactory receptor MOR261-6	1330	81

272

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2292	gi32052343	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4384160-4383228	1330	81
2292	gi9368991	Homo sapiens		1397	99
2293	gi29791964	Homo sapiens	Thrombospondin 4	2097	100
2293	gi311626	Homo sapiens	thrombospondin-4	2090	99
2293	gi4895079	Mus musculus	thrombospondin 4	2047	96
2294	gi24460119	Mus musculus	AF327451_1 JNK-associated leucine-zipper protein	6108	95
2294	gi24460121	Homo sapiens	AF327452_1 JNK-associated leucine-zipper protein	6282	98
2294	gi3116015	Homo sapiens	sperm specific protein	3848	100
2295	gi21654741	Homo sapiens	peptide/histidine transporter	2861	100
2295	gi2208839	Rattus norvegicus	peptide/histidine transporter	2484	87
2295	gi33126130	Homo sapiens	peptide/histidine transporter	2826	99
2296	gi19353264	Homo sapiens	Similar to dishevelled associated activator of morphogenesis 2	193	34
2296	gi2224703	Homo sapiens	KIAA0381	291	50
2296	gi30268369	Homo sapiens	hypothetical protein	291	50
2297	gi22760046	Homo sapiens	unnamed protein product	918	95
2297	gi27769120	Homo sapiens	Similar to hypothetical protein FLJ30921	918	95
2297	gi33417243	Mus musculus	B230312I18Rik protein	621	62
2298	gi12655913	Homo sapiens	AF227516_1 sprouty-4A	494	97
2298	gi12655915	Homo sapiens	AF227517_1 sprouty-4C	413	100
2298	gi29747900	Mus musculus	Sprouty homolog 4	347	83
2299	gi29692498	Mus musculus	NAAG-peptidase II	3438	87
2299	gi3211746	Sus scrofa	folylpoly-gamma-glutamate carboxypeptidase	2813	70
2299	gi4539525	Homo sapiens	NAALADase II protein	3872	99
2300	gi21750009	Homo sapiens	unnamed protein product	501	100
2300	gi23092685	Drosophila melanogaster	CG7020-PA	150	76
2300	gi23512248	Homo sapiens	Similar to DISCO Interacting Protein 2	238	56
2301	gi21410507	Mus musculus	Plxnb2 protein	465	75
2301	gi6010211	Homo sapiens	semaphorin receptor	225	47
2301	gi9885259	Homo sapiens	AF149019_1 plexin-B3	228	47
2302	gi11692802	Homo sapiens	AF320294_1 ABCG8	287	88
2302	gi15088540	Homo sapiens	AF324494_1 sterolin-2	287	88
2302	gi15146444	Homo sapiens	AF351824_1 sterolin-2	287	88
2303	gi12652851	Homo sapiens	AAH00178 potassium channel modulatory factor	1987	100
2303	gi26453336	Homo sapiens	FIGC1	1983	99
2303	gi7677058	Homo sapiens	AF155652_1 potassium channel modulatory factor	1983	99
2305	gi24430369	Mus musculus	MMAC8	280	47
2305	gi31338848	Mus musculus	MAIR-Ia	285	46
2305	gi31338850	Mus musculus	MAIR-Ib	280	47
2306	gi31414326	Homo sapiens	MHC class I antigen	1941	99
2306	gi33187148	Homo sapiens	HLA-A2	1941	99
2306	gi403144	Homo sapiens	MHC class I lymphocyte antigen	1941	99

273

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2307	gi21667214	Homo sapiens	AF465767_1 bactericidal/permeability-increasing protein-like 3	743	90
2307	gi32490539	Homo sapiens	RY2G5	191	29
2307	gi57732	Rattus rattus	potential ligand-binding protein	231	32
2308	gi21667214	Homo sapiens	AF465767_1 bactericidal/permeability-increasing protein-like 3	743	90
2308	gi32490539	Homo sapiens	RY2G5	191	29
2308	gi57732	Rattus rattus	potential ligand-binding protein	231	32
2309	gi21667214	Homo sapiens	AF465767_1 bactericidal/permeability-increasing protein-like 3	743	90
2309	gi32490539	Homo sapiens	RY2G5	191	29
2309	gi57732	Rattus rattus	potential ligand-binding protein	231	32
2310	gi21667214	Homo sapiens	AF465767_1 bactericidal/permeability-increasing protein-like 3	743	90
2310	gi32490539	Homo sapiens	RY2G5	191	29
2310	gi57732	Rattus rattus	potential ligand-binding protein	231	32
2311	gi13529158	Homo sapiens	AAH05349	1137	99
2311	gi529514	Sus scrofa	neuronal endocrine protein	1073	94
2311	gi7718079	Homo sapiens	neuroendocrine protein 7B2	1129	99
2312	gi15029903	Mus musculus	Similar to proline-rich protein BstNI subfamily 2	175	31
2312	gi31746553	Caenorhabditis elegans	Collagen protein 51	171	35
2312	gi32698037	Caenorhabditis elegans		174	33
2313	gi13543081	Mus musculus	claudin 6	822	70
2313	gi4128041	Homo sapiens	claudin-9 protein	1116	100
2313	gi4325296	Mus musculus	claudin-9	1078	95
2314	gi18676638	Homo sapiens	FLJ00218 protein	574	95
2314	gi4587895	Rattus norvegicus	AF072509_1 glutamate receptor interacting protein 2	667	84
2314	gi6601555	Rattus norvegicus	glutamate receptor interacting protein 2	667	84
2315	gi23496442	Rattus norvegicus	disabled-1	2807	96
2315	gi3288852	Homo sapiens	disabled-1	2865	99
2315	gi8118615	Homo sapiens	AF263547_1 disabled-1	2842	99
2316	gi16877456	Homo sapiens	AAH16974	493	100
2316	gi20810324	Homo sapiens		493	100
2316	gi26351033	Mus musculus	unnamed protein product	444	91
2317	gi15430703	Homo sapiens	AF362953_1 testis specific serine/threonine kinase 2	1854	99
2317	gi2738898	Mus musculus	protein kinase	1684	89
2317	gi33590489	Rattus norvegicus	serine/threonine kinase 22B	1755	92
2318	gi12963879	Homo sapiens	prostaglandin D synthase	998	100
2318	gi13543568	Homo sapiens	PTGDS protein	998	100
2318	gi189772	Homo sapiens	prostaglandin D2 synthase	998	100

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2319	gi14336718	Homo sapiens	AE006464_18 similar to HAGH	656	99
2319	gi14336766	Homo sapiens	AE006639_8 hydroxyacylglutathione hydrolase	339	47
2319	gi20988885	Mus musculus	2810014I23Rik protein	583	78
2320	gi13397835	Homo sapiens	annexin A13 isoform b	1245	98
2320	gi33980	Homo sapiens	intestine-specific annexin	1252	98
2320	gi757784	Canis familiaris	annexin XIIIb	1151	91
2321	gi204222	Rattus norvegicus	GABA transporter protein	2124	90
2321	gi21707908	Homo sapiens	, member 1	2132	99
2321	gi31658	Homo sapiens	GABA transporter	2117	99
2323	gi20381266	Homo sapiens	Glypican 2	602	90
2323	gi440127	Rattus norvegicus	cerebroglycan	548	81
2323	gi5911318	Homo sapiens	AF105267_1 glypican-6	265	47
2324	gi18676470	Homo sapiens	FLJ00132 protein	1361	100
2324	gi19344068	Mus musculus	2700038E08Rik protein	2403	74
2324	gi23274106	Mus musculus	2700038E08Rik protein	2403	74
2325	gi25396387	Homo sapiens	alpha 2,6-sialyltransferase	467	98
2325	gi27650880	Homo sapiens	beta-galactoside alpha-2,6-sialyltransferase	467	98
2325	gi452751	Gallus gallus	Gal beta 1,4 GlcNAc alpha 2,6-sialyltransferase	268	58
2326	gi13344995	Homo sapiens	Cat Eye Syndrome critical region protein isoform 1	2004	99
2326	gi13344997	Homo sapiens	Cat Eye Syndrome critical region protein isoform 2	2001	100
2326	gi27503696	Homo sapiens	Similar to cat eye syndrome chromosome region, candidate 5	2001	100
2327	gi13344995	Homo sapiens	Cat Eye Syndrome critical region protein isoform 1	2004	99
2327	gi13344997	Homo sapiens	Cat Eye Syndrome critical region protein isoform 2	2001	100
2327	gi27503696	Homo sapiens	Similar to cat eye syndrome chromosome region, candidate 5	2001	100
2328	gi202592	Rattus norvegicus	prealpha-2-macroglobulin	238	40
2328	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2328	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2329	gi202592	Rattus norvegicus	prealpha-2-macroglobulin	238	40
2329	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2329	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2330	gi202592	Rattus norvegicus	prealpha-2-macroglobulin	238	40
2330	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2330	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2331	gi202592	Rattus norvegicus	prealpha-2-macroglobulin	238	40
2331	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2331	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2332	gi202592	Rattus norvegicus	prealpha-2-macroglobulin	238	40
2332	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2332	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2333	gi14789873	Mus musculus	Es31 protein	508	70
2333	gi17512361	Mus musculus	esterase 31	508	70

275

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2333	gi29476863	Mus musculus	Similar to esterase 31	516	69
2334	gi19909128	Homo sapiens	AF489528_1 transforming growth factor-beta binding protein-1S	189	100
2334	gi207286	Rattus norvegicus	TGF-beta masking protein large subunit	179	90
2334	gi339548	Homo sapiens	transforming growth factor-beta 1 binding protein precursor	189	100
2336	gi1388158	Gallus gallus	myomesin	429	37
2336	gi31418212	Homo sapiens	Myomesin 2	439	36
2336	gi407097	Homo sapiens	165kD protein	439	36
2339	gi12655442	Homo sapiens	keratin associated protein 4.2	706	86
2339	gi12655460	Homo sapiens	keratin associated protein 4.12	732	86
2339	gi12655464	Homo sapiens	keratin associated protein 4.15	761	99
2340	gi12655442	Homo sapiens	keratin associated protein 4.2	706	86
2340	gi12655460	Homo sapiens	keratin associated protein 4.12	732	86
2340	gi12655464	Homo sapiens	keratin associated protein 4.15	761	99
2341	gi12655442	Homo sapiens	keratin associated protein 4.2	706	86
2341	gi12655460	Homo sapiens	keratin associated protein 4.12	732	86
2341	gi12655464	Homo sapiens	keratin associated protein 4.15	761	99
2342	gi15722084	Homo sapiens		1930	99
2342	gi434306	Homo sapiens	lysosomal acid lipase; sterol esterase	1288	63
2342	gi506431	Homo sapiens	lysosomal acid lipase	1288	63
2343	gi15722084	Homo sapiens		1930	99
2343	gi434306	Homo sapiens	lysosomal acid lipase; sterol esterase	1288	63
2343	gi506431	Homo sapiens	lysosomal acid lipase	1288	63
2344	gi20152322	Homo sapiens	putative G-protein coupled receptor	1570	100
2344	gi32526601	Homo sapiens	GPRC5D	1576	100
2344	gi8118040	Homo sapiens	AF209923_1 orphan G-protein coupled receptor	1570	100
2345	gi17224598	Homo sapiens	AF293615_1 blood dendritic cell antigen 2 protein	1147	95
2345	gi17225337	Homo sapiens	AF325459_1 dendritic lectin	1147	95
2345	gi17225339	Homo sapiens	AF325460_1 dendritic lectin b isoform	953	82
2346	gi17224598	Homo sapiens	AF293615_1 blood dendritic cell antigen 2 protein	1147	95
2346	gi17225337	Homo sapiens	AF325459_1 dendritic lectin	1147	95
2346	gi17225339	Homo sapiens	AF325460_1 dendritic lectin b isoform	953	82
2347	gi21929119	Homo sapiens	seven transmembrane helix receptor	1588	100
2347	gi2792016	Homo sapiens	olfactory receptor	1393	100
2347	gi4092819	Homo sapiens	BC319430_5	1386	100
2348	gi2589172	Rattus norvegicus	mucin Muc3	308	36
2348	gi28436742	Mus musculus	Muc3 protein	295	37
2348	gi5911169	Homo sapiens	AF147790_1 transmembrane mucin 12	719	81
2349	gi3549152	Homo sapiens	R29124_1	180	36
2349	gi8101840	Papio hamadryas	AF259559_1 carcinoembryonic antigen-	182	35

276

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			family cell adhesion molecule w; CEACAMw		
2349	gi8101856	Cercopithecus aethiops	AF259567_1 carcinoembryonic antigen-family cell adhesion molecule 1-1; CEACAM1	179	33
2350	gi27924102	Mus musculus	2310075M15Rik protein	944	68
2350	gi29436830	Mus musculus	2310075M15Rik protein	944	68
2350	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	940	67
2351	gi27924102	Mus musculus	2310075M15Rik protein	944	68
2351	gi29436830	Mus musculus	2310075M15Rik protein	944	68
2351	gi6273399	Homo sapiens	AF200348_1 melanoma-associated antigen MG50	940	67
2352	gi10435776	Homo sapiens	unnamed protein product	1132	99
2352	gi32451585	Homo sapiens		681	60
2352	gi7264653	Mus musculus	AF180470_1 Kiaa0575	694	62
2353	gi20219008	Chlamydomonas reinhardtii	AF394181_1 coiled-coil flagellar protein	280	29
2353	gi23497711	Plasmodium falciparum 3D7	AE014826_49 rhoptry protein, putative	149	25
2353	gi5457791	Pyrococcus abyssi	smc1 chromosome segregation protein	150	22
2354	gi12654511	Homo sapiens	Torsin family 3, member A	1438	100
2354	gi14043167	Homo sapiens	Torsin family 3, member A	1438	100
2354	gi15079904	Homo sapiens	Torsin family 3, member A	1438	100
2356	gi15076843	Homo sapiens	AF233450_1 pecanex-like protein 1	948	72
2356	gi18157547	Mus musculus	AF237953_1 pecanex-like 3	1325	98
2356	gi6650377	Mus musculus	AF096286_1 pecanex 1	948	71
2357	gi15076843	Homo sapiens	AF233450_1 pecanex-like protein 1	948	72
2357	gi18157547	Mus musculus	AF237953_1 pecanex-like 3	1325	98
2357	gi6650377	Mus musculus	AF096286_1 pecanex 1	948	71
2358	gi1872200	Homo sapiens	alternatively spliced product using exon 13A	298	72
2358	gi2580578	Homo sapiens	ubiquitous TPR motif, Y isoform	301	70
2358	gi8572229	Homo sapiens	ubiquitous TPR-motif protein Y isoform	301	70
2359	gi12043567	Homo sapiens	unc-93 related protein	1544	97
2359	gi17390915	Mus musculus	unc93 homolog B	1350	85
2359	gi23271746	Mus musculus	Unc93b protein	1350	85
2360	gi15990461	Homo sapiens	AAH15612 ring finger protein 25	2465	100
2360	gi18490513	Mus musculus	Rnf25 protein	1983	82
2360	gi29179411	Mus musculus	Ring finger protein 25	1988	82
2361	gi14714684	Mus musculus	2810423E13Rik protein	632	83
2361	gi33086578	Rattus norvegicus	Ab2-276	385	82
2361	gi7295255	Drosophila melanogaster	CG8596-PA	307	46
2362	gi16930383	Pan troglodytes	AF383169_1 leukocyte immunoglobulin-like receptor e	172	38
2362	gi32396010	Bos taurus	immunoglobulin A Fc receptor	179	33

277

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2362	gi6563042	Homo sapiens	AF109683_1 leukocyte-associated Ig-like receptor 1b	179	24
2363	gi16930383	Pan troglodytes	AF383169_1 leukocyte immunoglobulin-like receptor e	172	38
2363	gi32396010	Bos taurus	immunoglobulin A Fc receptor	179	33
2363	gi6563042	Homo sapiens	AF109683_1 leukocyte-associated Ig-like receptor 1b	179	24
2364	gi21595190	Mus musculus	2510001A17Rik protein	366	98
2364	gi21707128	Homo sapiens	Ran binding protein 11	370	100
2364	gi6650612	Homo sapiens	AF111109_1 Ran binding protein 11	370	100
2367	gi11493419	Homo sapiens	AF130117_15 PRO1367	128	51
2367	gi6690223	Homo sapiens	AF090928_1 PRO0470	118	50
2367	gi6855613	Homo sapiens	AF113685_1 PRO0974	154	51
2369	gi3002527	Homo sapiens	neuronal thread protein AD7c-NTP	404	48
2369	gi32486167	Homo sapiens	AD7C-NTP	404	48
2369	gi6650810	Homo sapiens	AF118094_21 PRO1902	258	64
2370	gi13278391	Mus musculus	RIKEN cDNA 9430015G10	595	71
2370	gi14250646	Homo sapiens	FLJ20584 protein	803	98
2370	gi7020791	Homo sapiens	unnamed protein product	834	99
2371	gi16588454	Homo sapiens	AF312374_1 AGTRAP protein	823	100
2371	gi16878260	Homo sapiens	AAH17328 Similar to angiotensin II, type I receptor-associated protein	776	95
2371	gi9621816	Homo sapiens	AF165187_1 ATRAP	822	99
2372	gi12330704	Mus musculus	AF333770_1 cell recognition molecule CASPR4	539	82
2372	gi17986216	Homo sapiens	AF333769_1 cell recognition molecule CASPR3	633	97
2372	gi21961652	Mus musculus	contactin associated protein 4	539	82
2373	gi12330704	Mus musculus	AF333770_1 cell recognition molecule CASPR4	539	82
2373	gi17986216	Homo sapiens	AF333769_1 cell recognition molecule CASPR3	633	97
2373	gi21961652	Mus musculus	contactin associated protein 4	539	82
2374	gi11041469	Macaca fascicularis	UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase	1116	63
2374	gi21552746	Homo sapiens	AF410457_1 putative polypeptide N-acetylgalactosaminyltransferase	1670	100
2374	gi21552969	Mus musculus	AF467979_1 Williams-Beuren syndrome critical region gene 17	1656	98
2375	gi16198335	Drosophila melanogaster	SD08329p	411	47
2375	gi23092707	Drosophila melanogaster	CG17090-PA	411	47
2375	gi23092708	Drosophila melanogaster	CG17090-PB	411	47
2377	gi14571502	Homo sapiens	calcium-promoted Ras inactivator	1022	81
2377	gi15680152	Homo sapiens	AAH14420	317	41
2377	gi4185294	Homo sapiens	rasGAP-activating-like protein	289	36
2379	gi15128105	Mus musculus	AF397008_1 nephronectin	737	82

278

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2379	gi15430246	Mus musculus	nephronectin short isoform	737	82
2379	gi15430248	Mus musculus	nephronectin long isoform	737	82
2380	gi16041675	Homo sapiens	AAH15704 joined to JAZF1	2131	99
2380	gi17862954	Drosophila melanogaster	SD04959p	311	31
2380	gi28839713	Homo sapiens	Similar to joined to JAZF1	363	81
2381	gi29387355	Xenopus laevis		263	28
2381	gi3242649	Rana catesbeiana	alpha 1 type I collagen	297	28
2381	gi4140029	Cynops pyrrhogaster	alpha 1 type I collagen	277	27
2382	gi32967231	Homo sapiens	TAF3	481	100
2382	gi32967237	Homo sapiens	TAF3.2	619	100
2382	gi32967243	Mus musculus	TAF3	390	82
2383	gi32967231	Homo sapiens	TAF3	481	100
2383	gi32967237	Homo sapiens	TAF3.2	619	100
2383	gi32967243	Mus musculus	TAF3	390	82
2384	gi10443967	Homo sapiens	AF268610_1 THEG protein	298	60
2384	gi20306274	Homo sapiens	testicular haploid expressed gene	298	60
2384	gi7416134	Homo sapiens	testis-specific gene	298	60
2385	gi18480746	Mus musculus	olfactory receptor MOR261-10	1336	80
2385	gi21928655	Homo sapiens	seven transmembrane helix receptor	1427	90
2385	gi32052225	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4341246-4340281	1336	80
2386	gi18480746	Mus musculus	olfactory receptor MOR261-10	1336	80
2386	gi21928655	Homo sapiens	seven transmembrane helix receptor	1427	90
2386	gi32052225	Mus musculus	olfactory receptor GA_x6K02T2P3E9-4341246-4340281	1336	80
2387	gi13937888	Homo sapiens	AAH07052 Similar to heterogeneous nuclear ribonucleoprotein C	196	97
2387	gi337455	Homo sapiens	hnRNP C2 protein	196	97
2387	gi4139188	Mus musculus	heterogeneous nuclear ribonucleoprotein C1/C2; hnRNP C1/C2	190	95
2388	gi190259	Homo sapiens	neuron-specific protein	335	100
2388	gi190261	Homo sapiens	21 kDa protein	335	100
2388	gi56877	Rattus norvegicus	reading frame 1	331	98
2389	gi14573319	Homo sapiens	AF334755_1 interleukin-1 HY2	818	100
2389	gi14573321	Homo sapiens	AF334756_1 interleukin-1 HY2	818	100
2389	gi18025344	Homo sapiens	interleukin-1 receptor antagonist-like FIL1 theta	804	98
2390	gi27694303	Homo sapiens	Similar to keratin, hair, acidic, 6	694	69
2390	gi3724099	Homo sapiens	type I hair keratin 1	692	69
2390	gi3724114	Homo sapiens	type I hair keratin 6	694	69
2391	gi32488718	Oryza sativa (japonica cultivar-group)	OSJNBa0088H09.19	121	41

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2393	gi14595019	Homo sapiens	keratin 6 irs	362	98
2393	gi27901522	Homo sapiens	keratin 6 irs3	361	98
2393	gi27901524	Homo sapiens	keratin 6 irs4	353	95
2394	gi11066090	Homo sapiens	AF195192_1 matrix metalloprotease MMP-27	507	100
2394	gi12006364	Tupaia belangeri	AF281673_1 matrix metalloproteinase-27	458	91
2394	gi3511149	Gallus gallus	matrix metalloproteinase	353	60
2395	gi11066090	Homo sapiens	AF195192_1 matrix metalloprotease MMP-27	507	100
2395	gi12006364	Tupaia belangeri	AF281673_1 matrix metalloproteinase-27	458	91
2395	gi3511149	Gallus gallus	matrix metalloproteinase	353	60
2396	gi24710913	Homo sapiens	suppressor of fused	2599	100
2396	gi5739507	Homo sapiens	AF175770_1 suppressor of fused	2594	99
2396	gi6689894	Homo sapiens	AF159447_1 Suppressor of Fused	2599	100
2397	gi20387087	Oncorhynchus mykiss	like-2	155	32
2397	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability-increasing protein-like 2	535	100
2397	gi28173296	Cyprinus carpio	bactericidal permeability-increasing protein/lipopolysaccharide-binding protein	161	36
2398	gi19526647	Homo sapiens	AF462348_1 oxidored-nitro domain-containing protein	2019	99
2398	gi28175624	Mus musculus	RIKEN cDNA 1810007P19 gene	1704	86
2398	gi7303522	Drosophila melanogaster	CG13178-PA	214	29
2399	gi19526647	Homo sapiens	AF462348_1 oxidored-nitro domain-containing protein	2019	99
2399	gi28175624	Mus musculus	RIKEN cDNA 1810007P19 gene	1704	86
2399	gi7303522	Drosophila melanogaster	CG13178-PA	214	29
2400	gi2072977	Homo sapiens	putative p150	151	100
2400	gi339771	Homo sapiens	ORF2	151	100
2400	gi339777	Homo sapiens	ORF2 contains a reverse transcriptase domain.	151	100
2402	gi11493483	Homo sapiens	AF130117_48 PRO2550	303	64
2402	gi7020440	Homo sapiens	unnamed protein product	310	57
2402	gi7770139	Homo sapiens	AF119917_13 PRO1722	289	60
2404	gi1403325	Homo sapiens	MACH-beta-1	122	92
2404	gi1403327	Homo sapiens	MACH-beta-2	122	92
2405	gi1799570	Rattus norvegicus	TIP120	6200	99
2405	gi29792160	Homo sapiens	TIP120 protein	6213	99
2405	gi7688703	Homo sapiens	AF157326_1 TIP120 protein	6200	99
2406	gi13016701	Homo sapiens	activating coreceptor NKp80	1209	97
2406	gi22449867	Macaca fascicularis	NKp80 NK receptor	1105	87
2406	gi7188567	Homo sapiens	AF175206_1 lectin-like receptor F1	1209	97

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2408	gi21619190	Homo sapiens	-like 1X-linked	233	80
2408	gi27695407	Mus musculus	Tbl1x protein	233	80
2408	gi30353941	Homo sapiens	TBL1X protein	233	80
2409	gi12804613	Homo sapiens	AAH01728	670	82
2409	gi13279113	Homo sapiens	AAH04281	670	82
2409	gi14043598	Homo sapiens	AAH07776	670	82
2410	gi12804613	Homo sapiens	AAH01728	670	82
2410	gi13279113	Homo sapiens	AAH04281	670	82
2410	gi14043598	Homo sapiens	AAH07776	670	82
2411	gi12804613	Homo sapiens	AAH01728	670	82
2411	gi13279113	Homo sapiens	AAH04281	670	82
2411	gi14043598	Homo sapiens	AAH07776	670	82
2412	gi13182755	Homo sapiens	AF212237_1 HPHRP	1816	99
2412	gi15929309	Homo sapiens	Phosphotriesterase related	1824	100
2412	gi29791939	Homo sapiens	phosphotriesterase related	1824	100
2414	gi22539701	Mus musculus	4930506M07Rik protein	2153	93
2414	gi4778	Saccharomyces cerevisiae	Uso1 protein	215	23
2414	gi677198	Saccharomyces cerevisiae	putative	217	23
2415	gi27899969	Homo sapiens	unnamed protein product	208	66
2415	gi27900262	Homo sapiens	unnamed protein product	208	66
2415	gi6690248	Homo sapiens	AF090942_1 PRO0657	192	57
2419	gi13377880	Cricetulus longicaudatus	AF336043_1 arginine N-methyltransferase p82 isoform	2585	85
2419	gi13377882	Cricetulus longicaudatus	AF336044_1 arginine N-methyltransferase p77 isoform	2534	86
2419	gi13879453	Mus musculus	cDNA sequence BC006705	2565	87
2420	gi16306618	Homo sapiens	AAH01482 phosphatidylserine decarboxylase	1645	99
2420	gi191185	Cricetulus griseus	phosphatidylserine decarboxylase	1544	93
2420	gi27371042	Xenopus laevis	Similar to phosphatidylserine decarboxylase	958	57
2421	gi30041	Homo sapiens	COL2A1	122	28
2421	gi450394	Homo sapiens	alpha-1 type II collagen	122	28
2421	gi930050	Homo sapiens		122	28
2422	gi13874437	Homo sapiens	cerebral protein-11	159	75
2422	gi20987344	Mus musculus	LOC212904 protein	618	69
2422	gi24980850	Homo sapiens		765	100
2423	gi13543940	Homo sapiens	Hypothetical protein DKFZp434B195	2094	99
2423	gi14035978	Homo sapiens	unnamed protein product	2080	98
2423	gi16923351	Homo sapiens	AF204270_1 RbBP-35	1419	98
2424	gi18676660	Homo sapiens	FLJ00229 protein	665	99
2424	gi25955706	Homo sapiens	Similar to hypothetical protein MGC38041	665	99
2424	gi32484169	Homo sapiens		665	99
2425	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	410	89
2425	gi29293087	Homo sapiens	dipeptidyl peptidase 9	410	89
2425	gi3513303	Homo sapiens	R26984_1	476	100
2426	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	410	89

281

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2426	gi29293087	Homo sapiens	dipeptidyl peptidase 9	410	89
2426	gi3513303	Homo sapiens	R26984_1	476	100
2427	gi27549552	Homo sapiens	dipeptidyl peptidase IV-related protein-2	410	89
2427	gi29293087	Homo sapiens	dipeptidyl peptidase 9	410	89
2427	gi3513303	Homo sapiens	R26984_1	476	100
2428	gi13097642	Homo sapiens	Ribosomal protein S25	169	100
2428	gi13279149	Homo sapiens	Ribosomal protein S25	169	100
2428	gi13436422	Homo sapiens	Ribosomal protein S25	169	100
2429	gi21756739	Homo sapiens	unnamed protein product	2539	96
2429	gi23270822	Homo sapiens		2427	96
2429	gi6453538	Homo sapiens	hypothetical protein	2061	99
2430	gi12652695	Homo sapiens	AAH00096 HtrA-like serine protease	1611	92
2430	gi5870865	Homo sapiens	serine protease	1611	92
2430	gi7672669	Homo sapiens	AF141305_1 serine protease Htra2	1611	92
2431	gi24078514	Mus musculus	AF454954_1 crossveinless-2	561	95
2431	gi32816043	Mus musculus	BMP-binding endothelial regulator precursor protein	561	95
2431	gi32892146	Homo sapiens	crossveinless-2	595	100
2432	gi16502169	Salmonella enterica subsp. enterica serovar Typhi	putative DNA methylase	756	85
2432	gi29137981	Salmonella enterica subsp. enterica serovar Typhi Ty2	putative DNA methylase	756	85
2432	gi498768	Serratia marcescens	Deoxyadenosyl-methyltransferase	337	47
2433	gi16974751	Gallus gallus	CALII	184	44
2433	gi19908346	Gallus gallus	chondrogenesis associated lipocalin	137	37
2433	gi22090638	Gallus gallus	lipocalin-type prostaglandin D synthase	137	37
2434	gi17132791	Nostoc sp. PCC 7120	asparaginyl-tRNA synthetase	766	44
2434	gi22296200	Thermosynechococcus elongatus BP-1	asparaginyl-tRNA synthetase	767	41
2434	gi30259286	Bacillus anthracis str. Ames	asparaginyl-tRNA synthetase	774	43
2435	gi12655061	Homo sapiens	AAH01380	532	88
2435	gi23574788	Macaca fascicularis	succinate dehydrogenase flavoprotein subunit	539	89
2435	gi5759173	Homo sapiens	succinate dehydrogenase flavoprotein subunit	532	88
2436	gi21928188	Mus musculus	GPI-gamma 4; GPIgamma4	853	67
2436	gi29747988	Mus musculus	GPI-gamma 4	853	67
2436	gi30931171	Mus musculus	GPIgamma4 protein	853	67
2437	gi15082311	Homo sapiens	AAH12061 -binding protein 3	631	98
2437	gi27503479	Mus musculus	Pcbp3 protein	631	98
2437	gi9957165	Homo sapiens	AF176329_1 alphaCP-3	631	98
2438	gi16553246	Homo sapiens	unnamed protein product	254	98
2438	gi21739662	Homo sapiens	hypothetical protein	218	88
2438	gi21752375	Homo sapiens	unnamed protein product	218	88
2439	gi12804943	Homo sapiens	AAH01924 beta	1660	90

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2439	gi189762	Homo sapiens	pyruvate dehydrogenase E1-beta subunit	1663	91
2439	gi190792	Homo sapiens	pyruvate dehydrogenase E1-beta subunit precursor	1663	91
2440	gi164851	Oryctolagus cuniculus	calsequestrin precursor	1903	92
2440	gi2618621	Mus musculus	skeletal muscle calsequestrin	1921	93
2440	gi688292	Homo sapiens	calmitine; calsequestrine	2012	99
2441	gi1177622	Saccharomyces cerevisiae	AOF1001	177	30
2441	gi13592175	Leishmania major	AC084329_1 ppg3	193	26
2441	gi28828184	Dictyostelium discoideum	similar to Leishmania major. Ppg3	192	26
2442	gi20380863	Homo sapiens	Similar to T cell receptor beta locus	1364	84
2442	gi307487	Homo sapiens	T-cell receptor beta	1498	93
2442	gi8515902	Homo sapiens	T cell receptor beta chain	1300	84
2444	gi14599484	Homo sapiens	AF333952_1 small proline-rich protein 2B	453	98
2444	gi3367693	Homo sapiens	small proline-rich protein	458	100
2444	gi385227	Homo sapiens	small proline-rich protein 2	453	98
2445	gi13876336	Mus musculus	protocadherin gamma A5	4081	84
2445	gi5456942	Homo sapiens	protocadherin gamma A5	4744	99
2445	gi5457072	Homo sapiens	AF152512_1 protocadherin gamma A5 short form protein	4109	100
2447	gi200962	Mus musculus	serine 1 ultra high sulfur protein	262	45
2447	gi200964	Mus musculus	serine 2 ultra high sulfur protein	296	49
2447	gi3228237	Homo sapiens	ultra high sulfur keratin	261	48
2448	gi14764499	Homo sapiens	zinc finger protein	849	66
2448	gi1504006	Homo sapiens	similar to human ZFY protein.	442	36
2448	gi28204954	Mus musculus	Similar to zinc finger protein	771	70
2450	gi17223709	Homo sapiens	selenoprotein SelM	235	100
2450	gi17223711	Mus musculus	selenoprotein SelM	188	78
2450	gi26351995	Mus musculus	unnamed protein product	162	76
2451	gi28848644	Homo sapiens	p02 protein	181	100
2451	gi30354510	Homo sapiens	TPT1 protein	181	100
2451	gi33285832	Homo sapiens	TCTP	181	100
2452	gi13937829	Homo sapiens	AAH07016	946	100
2452	gi18606299	Homo sapiens		946	100
2452	gi3360432	Homo sapiens	osteopontin	946	100
2453	gi14326586	Homo sapiens	AF386078_1 serine-cysteine proteinase inhibitor clade C member 1	360	92
2453	gi179130	Homo sapiens	antithrombin III	360	92
2453	gi18490839	Homo sapiens	, member 1	360	92
2454	gi37231	Homo sapiens	DNA topoisomerase II	8439	99
2454	gi3869382	Homo sapiens	DNA topoisomerase II beta	8299	99
2454	gi790988	Cricetulus longicaudatus		8167	96
2455	gi1881713	Rattus norvegicus	fatty acid transport protein	222	84
2455	gi20810561	Mus musculus	, member 1	219	82
2455	gi563829	Mus musculus	fatty acid transport protein	219	82

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2456	gi13277626	Mus musculus	homolog, subunit 7a	247	57
2456	gi15215085	Mus musculus	Cops7b protein	428	98
2456	gi3309176	Mus musculus	COP9 complex subunit 7b	428	98
2457	gi180251	Homo sapiens	precerebellin	183	48
2457	gi6942096	Mus musculus	CBLN3	472	90
2457	gi6942098	Mus musculus	AF218380_1 CBLN3	472	90
2458	gi17861952	Drosophila melanogaster	LD01947p	196	55
2458	gi31432182	Oryza sativa (japonica cultivar-group)	putative RIM2 protein	158	42
2458	gi7291183	Drosophila melanogaster	CG1826-PA	196	55
2459	gi20387087	Oncorhynchus mykiss	like-2	155	32
2459	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability-increasing protein-like 2	535	100
2459	gi28173296	Cyprinus carpio	bactericidal permeability-increasing protein/lipopolysaccharide-binding protein	161	36
2460	gi20387087	Oncorhynchus mykiss	like-2	155	32
2460	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability-increasing protein-like 2	535	100
2460	gi28173296	Cyprinus carpio	bactericidal permeability-increasing protein/lipopolysaccharide-binding protein	161	36
2461	gi20387087	Oncorhynchus mykiss	like-2	155	32
2461	gi21667212	Homo sapiens	AF465766_1 bactericidal/permeability-increasing protein-like 2	535	100
2461	gi28173296	Cyprinus carpio	bactericidal permeability-increasing protein/lipopolysaccharide-binding protein	161	36
2462	gi10435038	Homo sapiens	unnamed protein product	1718	96
2462	gi18257341	Mus musculus	Expressed sequence AW060207	1044	63
2462	gi24659229	Homo sapiens	hypothetical protein FLJ13150	1727	97
2464	gi27469556	Homo sapiens	Putative neuronal cell adhesion molecule	180	94
2464	gi4206390	Homo sapiens	putative neuronal cell adhesion molecule	180	94
2465	gi12667401	Homo sapiens	AF326731_1 NUF2R	2336	99
2465	gi14317902	Homo sapiens	kinetochore protein Nuf2	2336	99
2465	gi18043223	Mus musculus	NUF2R protein	1744	72
2466	gi23321257	Homo sapiens	ezrin-binding partner PACE-1	3482	97
2466	gi24209887	Homo sapiens	ezrin-binding protein PACE-1	3381	90
2466	gi29144929	Mus musculus	Ezrin-binding partner PACE-1	2738	75
2467	gi21634823	Homo sapiens	AF389428_1 semaphorin 6D	1487	97

284

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
			isoform 3		
2467	gi21634825	Homo sapiens	AF389429_1 semaphorin 6D isoform 4	1487	97
2467	gi21634827	Homo sapiens	AF389430_1 semaphorin 6D isoform 1	1487	97
2468	gi13543141	Mus musculus	Slc37a3 protein	141	52
2469	gi21671105	Homo sapiens	RAD52B	511	100
2469	gi23468352	Homo sapiens	Similar to RAD52B	511	100
2469	gi32967621	Mus musculus	2410008M22Rik protein	311	66
2470	gi28626251	Homo sapiens	calcium-permeable store-operated channel TRPM3c	289	91
2470	gi28626253	Homo sapiens	calcium-permeable store-operated channel TRPM3d	289	91
2470	gi28626255	Homo sapiens	calcium-permeable store-operated channel TRPM3e	289	91
2472	gi20987880	Mus musculus	E130103I17Rik protein	1605	71
2472	gi28204917	Mus musculus	E130103I17Rik protein	1594	71
2472	gi4588087	Homo sapiens	AF095771_1 PTH-responsive osteosarcoma B1 protein	1864	89
2473	gi13591434	Homo sapiens		413	74
2473	gi13591435	Homo sapiens		416	87
2473	gi19913471	Homo sapiens		413	74
2474	gi28372402	Homo sapiens	truncated transmembrane transport protein	1271	100
2474	gi31324239	Homo sapiens	proton-coupled amino acid transporter	1263	100
2474	gi31871291	Homo sapiens	proton/amino acid transporter 1	1263	100
2475	gi28372402	Homo sapiens	truncated transmembrane transport protein	1271	100
2475	gi31324239	Homo sapiens	proton-coupled amino acid transporter	1263	100
2475	gi31871291	Homo sapiens	proton/amino acid transporter 1	1263	100
2476	gi11138040	Homo sapiens	rat myomegalin mRNA is reported in Acc# AF139185~similar to rat myomegalin	828	97
2476	gi11138042	Homo sapiens	rat myomegalin mRNA is reported in Acc# AF139185~similar to rat myomegalin	1091	93
2476	gi19263586	Homo sapiens	similar to rat myomegalin	1085	93
2477	gi19263005	Ciona intestinalis	leucine-rich repeat dynein light chain	367	66
2477	gi2760161	Anthocidaris crassispina	outer arm dynein light chain 2	338	63
2477	gi7303901	Drosophila melanogaster	CG8800-PA	265	51
2478	gi12666531	Homo sapiens	putative b,b-carotene-9',10'-dioxygenase	917	99
2478	gi14582265	Homo sapiens	AF276432_1 putative carotene dioxygenase	930	100
2478	gi27370671	Homo sapiens	Similar to beta-carotene dioxygenase 2	930	100
2479	gi12666531	Homo sapiens	putative b,b-carotene-9',10'-dioxygenase	917	99

285

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2479	gi14582265	Homo sapiens	AF276432_1 putative carotene dioxygenase	930	100
2479	gi27370671	Homo sapiens	Similar to beta-carotene dioxygenase 2	930	100
2480	gi1079734	Mus musculus	citron	718	97
2480	gi30088970	Homo sapiens	rho/rac-interacting citron kinase	696	99
2480	gi3599509	Mus musculus	rho/rac-interacting citron kinase	689	97
2481	gi24980821	Homo sapiens	box polypeptide 26	258	100
2481	gi32485107	Homo sapiens	nexin-related serine protease inhibitor	731	94
2481	gi6062874	Homo sapiens	candidate tumor suppressor protein DICE1	258	100
2482	gi13383364	Homo sapiens	claudin-1	1095	99
2482	gi15214678	Homo sapiens	AAH12471 claudin 1	1095	99
2482	gi7381083	Homo sapiens	AF134160_1 claudin-1	1095	99
2483	gi22902436	Mus musculus	Sphingosine-1-phosphate phosphatase 1	616	40
2483	gi23345324	Homo sapiens	sphingosine 1-phosphate phosphohydrolase 2	1513	99
2483	gi29436890	Mus musculus	Similar to sphingosine-1-phosphate phosphatase 2	1406	90
2484	gi2072977	Homo sapiens	putative p150	137	79
2484	gi339771	Homo sapiens	ORF2	137	79
2484	gi339777	Homo sapiens	ORF2 contains a reverse transcriptase domain.	137	79
2485	gi2072977	Homo sapiens	putative p150	137	79
2485	gi339771	Homo sapiens	ORF2	137	79
2485	gi339777	Homo sapiens	ORF2 contains a reverse transcriptase domain.	137	79
2487	gi18033185	Danio rerio	AF330001_1 UNC45-related protein	1491	79
2487	gi27436424	Mus musculus	striated muscle UNC45	1757	95
2487	gi27436426	Homo sapiens	striated muscle UNC45	1800	98
2488	gi26801168	Gallus gallus	condensin complex subunit	1330	44
2488	gi3851586	Homo sapiens	chromosome-associated protein-C	1123	63
2488	gi4092846	Homo sapiens	chromosome-associated polypeptide-C	1123	63
2489	gi2407911	Homo sapiens	CO16	1252	99
2489	gi29437323	Mus musculus	Similar to cDNA for differentially expressed CO16 gene	226	40
2489	gi6013073	Mus musculus	HemT-3 protein	141	27
2490	gi13157560	Homo sapiens		2246	99
2490	gi18147612	Homo sapiens	metalloprotease disintegrin	2246	99
2490	gi21908030	Homo sapiens	a disintegrin and metalloprotease domain 33	2230	98
2491	gi15145793	Sus scrofa	basic proline-rich protein	186	34
2491	gi3858883	Acanthamoeba castellanii	myosin I heavy chain kinase	218	37
2491	gi4206769	Acanthamoeba castellanii	myosin I heavy chain kinase	218	37
2492	gi1136434	Homo sapiens	KIAA0187	198	72

286

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2492	gi21410151	Mus musculus	LOC213895 protein	173	62
2492	gi27696627	Homo sapiens	Ribosome biogenesis protein BMS1 homolog	198	72
2493	gi13559063	Homo sapiens		747	100
2493	gi24416538	Mus musculus	1700001D09Rik protein	631	72
2493	gi9963863	Homo sapiens	AF226731_1 AD026	688	99
2495	gi156258	Caenorhabditis elegans	collagen	139	33
2495	gi21105301	Mytilus galloprovincialis	AF448525_1 precollagen-P	152	28
2495	gi2388676	Mytilus edulis	precollagen P	148	29
2496	gi156258	Caenorhabditis elegans	collagen	139	33
2496	gi21105301	Mytilus galloprovincialis	AF448525_1 precollagen-P	152	28
2496	gi2388676	Mytilus edulis	precollagen P	148	29
2497	gi156258	Caenorhabditis elegans	collagen	139	33
2497	gi21105301	Mytilus galloprovincialis	AF448525_1 precollagen-P	152	28
2497	gi2388676	Mytilus edulis	precollagen P	148	29
2498	gi20380052	Homo sapiens		372	32
2498	gi20380522	Mus musculus	Col3a1 protein	368	31
2498	gi29144943	Mus musculus	Col3a1 protein	368	31
2499	gi14035874	Homo sapiens	unnamed protein product	1100	99
2499	gi14035876	Homo sapiens	unnamed protein product	1043	99
2499	gi20070842	Homo sapiens	similar to hypothetical protein FLJ13448	1297	99
2501	gi2072964	Homo sapiens	putative p150	399	81
2501	gi2072967	Homo sapiens	putative p150	400	81
2501	gi339777	Homo sapiens	ORF2 contains a reverse transcriptase domain.	399	81
2502	gi30040280	Shigella flexneri 2a str. 2457T	IS103 orf	731	98
2502	gi30041139	Shigella flexneri 2a str. 2457T	IS103 orf	731	98
2502	gi466695	Escherichia coli	orfA in IS150	731	98
2503	gi12698037	Homo sapiens	KIAA1746 protein	341	100
2503	gi26344121	Mus musculus	unnamed protein product	318	92
2503	gi26351415	Mus musculus	unnamed protein product	318	92
2504	gi20269073	Homo sapiens	putative lipid kinase	1035	99
2504	gi21624340	Homo sapiens	ceramide kinase	1035	99
2504	gi21624342	Mus musculus	ceramide kinases	829	81
2505	gi312584	Mus musculus	biliary glycoprotein	165	27
2505	gi312586	Mus musculus	biliary glycoprotein	165	27
2505	gi312590	Mus musculus	biliary glycoprotein	174	30
2506	gi312584	Mus musculus	biliary glycoprotein	165	27
2506	gi312586	Mus musculus	biliary glycoprotein	165	27
2506	gi312590	Mus musculus	biliary glycoprotein	174	30
2507	gi1480744	Equus caballus	type II collagen	346	29
2507	gi30041	Homo sapiens	COL2A1	344	29
2507	gi450394	Homo sapiens	alpha-1 type II collagen	344	29
2508	gi1483580	Rattus norvegicus	NTR2 receptor	911	81
2508	gi18490912	Homo sapiens	neurotensin receptor 2	1072	95

287

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2508	gi3901028	Homo sapiens	neurotensin receptor 2	1074	95
2509	gi1049104	Homo sapiens	dystonin isoform 1	221	100
2509	gi14530942	Homo sapiens	dystonin 2	221	100
2509	gi14530944	Homo sapiens	dystonin 2	221	100
2510	gi1049104	Homo sapiens	dystonin isoform 1	221	100
2510	gi14530942	Homo sapiens	dystonin 2	221	100
2510	gi14530944	Homo sapiens	dystonin 2	221	100
2512	gi1572721	Homo sapiens	megakaryocyte stimulating factor; MSF	203	23
2512	gi16041156	Macaca fascicularis	X-ray radiation resistance associated 1 protein	710	66
2512	gi18676652	Homo sapiens	FLJ00225 protein	761	70
2513	gi1572721	Homo sapiens	megakaryocyte stimulating factor; MSF	203	23
2513	gi16041156	Macaca fascicularis	X-ray radiation resistance associated 1 protein	710	66
2513	gi18676652	Homo sapiens	FLJ00225 protein	761	70
2514	gi26346328	Mus musculus	unnamed protein product	965	93
2514	gi33417011	Mus musculus		965	93
2514	gi6330169	Homo sapiens	KIAA1164 protein	1005	99
2515	gi26346328	Mus musculus	unnamed protein product	965	93
2515	gi33417011	Mus musculus		965	93
2515	gi6330169	Homo sapiens	KIAA1164 protein	1005	99
2516	gi12857668	Mus musculus	unnamed protein product	123	43
2516	gi26327823	Mus musculus	unnamed protein product	123	43
2517	gi17429038	Ralstonia solanacearum	PROBABLE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE PROTEIN	676	61
2517	gi22776354	Oceanobacillus iheyensis HTE831	acyl-CoA dehydrogenase	660	63
2517	gi28280023	Mus musculus	5730439E10Rik protein	974	84
2518	gi17429038	Ralstonia solanacearum	PROBABLE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE PROTEIN	676	61
2518	gi22776354	Oceanobacillus iheyensis HTE831	acyl-CoA dehydrogenase	660	63
2518	gi28280023	Mus musculus	5730439E10Rik protein	974	84
2519	gi19070124	Mus musculus	AF233346_1 zinc transporter-like 3 protein	895	95
2519	gi20563194	Mus musculus	AF395840_1 zinc transporter 6	883	93
2519	gi33338012	Homo sapiens	AF173387_1 MSTP103	759	94
2520	gi212451	Gallus gallus	nonmuscle myosin heavy chain	182	20
2520	gi212452	Gallus gallus	nonmuscle myosin heavy chain	182	20
2520	gi4115748	Bos taurus	nonmuscle myosin heavy chain B	182	19
2521	gi18605758	Mus musculus	9030409G11Rik protein	1257	94
2521	gi6526769	Homo sapiens	HRIHFB2003	1200	96
2521	gi7291408	Drosophila melanogaster	CG11206-PA	263	26
2524	gi13182757	Homo sapiens	AF212238_1 HTPAP	843	100
2524	gi21542541	Homo sapiens	Similar to HTPAP protein	808	100
2524	gi28381093	Drosophila	CG12746-PD	410	50

288

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
		melanogaster			
2525	gi13182757	Homo sapiens	AF212238_1 HTPAP	843	100
2525	gi21542541	Homo sapiens	Similar to HTPAP protein	808	100
2525	gi28381093	Drosophila melanogaster	CG12746-PD	410	50
2527	gi16416764	Homo sapiens	AF315594_1 FKSG16	1027	100
2527	gi19353603	Mus musculus	D11Ert18e protein	337	41
2527	gi31873637	Homo sapiens	hypothetical protein	1014	100
2528	gi16416764	Homo sapiens	AF315594_1 FKSG16	1027	100
2528	gi19353603	Mus musculus	D11Ert18e protein	337	41
2528	gi31873637	Homo sapiens	hypothetical protein	1014	100
2529	gi32330803	Mus musculus	podocan protein	1095	90
2529	gi32330805	Homo sapiens	podocan protein	1205	97
2529	gi3786312	Homo sapiens	extracellular matrix protein	281	33
2530	gi20258604	Homo sapiens	sialic acid binding Ig-like lectin 5	2913	99
2530	gi2411475	Homo sapiens	OB binding protein-2	2913	99
2530	gi9454520	Homo sapiens	AC018755_5 SIGLEC5	2913	99
2531	gi20258604	Homo sapiens	sialic acid binding Ig-like lectin 5	2913	99
2531	gi2411475	Homo sapiens	OB binding protein-2	2913	99
2531	gi9454520	Homo sapiens	AC018755_5 SIGLEC5	2913	99
2532	gi13183078	Homo sapiens	AF237652_1 a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3	602	74
2532	gi15099921	Homo sapiens	AF176313_1 ADAM-TS related protein 1	874	98
2532	gi20987759	Homo sapiens	Similar to ADAMTS-like 1	886	99
2533	gi178836	Homo sapiens	apolipoprotein C-II	506	100
2533	gi30582255	Homo sapiens	apolipoprotein C-II	500	99
2533	gi757915	Homo sapiens	apoCII protein	506	100
2534	gi178836	Homo sapiens	apolipoprotein C-II	506	100
2534	gi30582255	Homo sapiens	apolipoprotein C-II	500	99
2534	gi757915	Homo sapiens	apoCII protein	506	100
2536	gi17389292	Homo sapiens	LDL induced EC protein	914	98
2536	gi5924319	Homo sapiens	AF184939_1 LDL induced EC protein	914	98
2536	gi8518179	Homo sapiens	LDL induced endothelial cell protein	941	76
2537	gi28974490	Homo sapiens	lipoma HMGIC fusion-partner-like protein	1071	100
2537	gi30102428	Rattus norvegicus	HMGIC fusion-partner-like protein	1038	95
2537	gi30411045	Mus musculus	Similar to lipoma HMGIC fusion partner	1037	94
2538	gi14603353	Homo sapiens	AAH10130 CGI-43 protein	2362	94
2538	gi23092946	Drosophila melanogaster	CG14980-PB	537	28
2538	gi4929555	Homo sapiens	AF151801_1 CGI-43 protein	2219	89
2539	gi12654633	Homo sapiens	Protein inhibitor of activated STAT3	179	84
2539	gi18606318	Mus musculus	Protein inhibitor of activated STAT 3, isoform 1	179	84

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2539	gi30582911	Homo sapiens	protein inhibitor of activated STAT3	179	84
2540	gi27449075	Oreochromis mossambicus	stearoyl-CoA desaturase	743	69
2540	gi29294686	Homo sapiens	SCD4 protein	737	100
2540	gi30350098	Homo sapiens	AF389338_1 acyl-CoA-desaturase	1016	100
2541	gi1000867	Homo sapiens	DNA mismatch repair protein	1931	100
2541	gi1000869	Homo sapiens	DNA mismatch repair protein	1931	100
2541	gi18204306	Homo sapiens	AAH21566	1931	100
2542	gi11862941	Mus musculus	DDM36E	430	48
2542	gi19570398	Homo sapiens	hDDM36	439	49
2542	gi7650186	Mus musculus	AF176694_1 neighbor of Punc e11 protein	430	48
2543	gi21744725	Homo sapiens	AF478693_1 glycosyl-phosphatidyl-inositol-MAM	717	97
2543	gi25005318	Sus scrofa	MAM domain containing glycosylphosphatidylinositol anchor 1	672	91
2543	gi25005320	Sus scrofa	glycosylphosphatidylinositol anchor 1 protein	672	91
2544	gi12276198	Homo sapiens	AF333487_1 FKSG40	543	96
2544	gi12408250	Homo sapiens	FKSG28	543	96
2544	gi18652934	Xenopus laevis	Mig30	514	48
2545	gi16769552	Drosophila melanogaster	LD38375p	367	51
2545	gi27696627	Homo sapiens	Ribosome biogenesis protein BMS1 homolog	684	93
2545	gi7294027	Drosophila melanogaster	CG7728-PA	367	51
2546	gi12842044	Mus musculus	unnamed protein product	375	72
2546	gi18921437	Mus musculus	2010004A03Rik protein	375	72
2546	gi20987450	Homo sapiens	LOC146433	468	91
2547	gi1016012	Rattus norvegicus	neural cell adhesion protein BIG-2 precursor	543	93
2547	gi26891535	Homo sapiens	contactin 4	570	100
2547	gi29837411	Homo sapiens	BIG-2	570	100
2548	gi30102449	Homo sapiens	lipoma HMGIC fusion-partner-like protein	822	100
2548	gi30908798	Homo sapiens	lipoma HMGIC fusion partner-like protein 4	676	78
2548	gi30908800	Rattus norvegicus	lipoma HMGIC fusion partner-like protein 4	675	78
2549	gi13097705	Homo sapiens	AAH03559 , member 3	237	52
2549	gi1340142	Homo sapiens	alpha1-antichymotrypsin	237	52
2549	gi4165890	Homo sapiens	alpha-1-antichymotrypsin precursor	237	52
2550	gi1850850	Murid herpesvirus 4	serine threonine rich glycoprotein	207	33
2550	gi21618556	Homo sapiens		4040	97
2550	gi33304372	Homo sapiens	tastin	4035	97
2551	gi12053849	Homo sapiens	DREV protein	1649	98
2551	gi12053851	Homo sapiens	DREV1 protein	1633	98
2551	gi12053853	Homo sapiens	DREV protein	1649	98

290

TABLE 2 B

SEQ_ID	Hit_ID	Species	Description	S_score	Percentage_Identity
2553	gi11990779	Homo sapiens		273	50
2553	gi22760096	Homo sapiens	unnamed protein product	538	100
2553	gi28279813	Homo sapiens	Similar to hypothetical protein DKFZp434A171	515	97
2554	gi11125348	Homo sapiens	putative protein kinase	2419	99
2554	gi6933864	Homo sapiens	kinase deficient protein KDP	2419	99
2554	gi8272557	Rattus norvegicus	AF227741_1 protein kinase WNK1	2340	96
2555	gi11125348	Homo sapiens	putative protein kinase	2419	99
2555	gi6933864	Homo sapiens	kinase deficient protein KDP	2419	99
2555	gi8272557	Rattus norvegicus	AF227741_1 protein kinase WNK1	2340	96
2556	gi3599339	Mus musculus domesticus	ORF2	138	60
2556	gi3599342	Mus musculus domesticus	ORF2	138	60
2556	gi3599347	Mus musculus domesticus	ORF2	138	60
2557	gi15020809	Takifugu rubripes	putative methionyl tRNA synthetase	674	74
2557	gi17861592	Drosophila melanogaster	GH13807p	567	61
2557	gi23171238	Drosophila melanogaster	CG31322-PA	567	61
2558	gi15341975	Homo sapiens	AAH13184 Similar to major histocompatibility complex, class II, DP beta 1	432	72
2558	gi17389919	Homo sapiens	AAH17967 Similar to major histocompatibility complex, class II, DP beta 1	814	100
2558	gi188479	Homo sapiens	HLA-DPB1	432	72
2559	gi15779083	Homo sapiens	AAH14609	1122	90
2559	gi3342737	Homo sapiens	R26660_2, partial CDS	967	86
2559	gi3478640	Homo sapiens	R26660_2, partial CDS	138	89
2560	gi15779083	Homo sapiens	AAH14609	1122	90
2560	gi3342737	Homo sapiens	R26660_2, partial CDS	967	86
2560	gi3478640	Homo sapiens	R26660_2, partial CDS	138	89
2561	gi13991167	Homo sapiens	sialic acid-binding immunoglobulin-like lectin-like long splice variant	661	99
2561	gi14625822	Homo sapiens	AF282256_1 Siglec-L1	661	99
2561	gi23272769	Homo sapiens	SIGLEC-like 1	661	99
2562	gi15132186	Homo sapiens	unnamed protein product	1122	88
2562	gi15132529	Homo sapiens	unnamed protein product	1122	88
2562	gi21439502	Homo sapiens	unnamed protein product	1122	88
2563	gi202592	Rattus norvegicus	prealpha-2-macroglobulin	238	40
2563	gi671864	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2563	gi671865	Gallus gallus	ovomacroglobulin, ovostatin	230	40
2564	gi25990364	Homo sapiens	AF319622_1 P-glycoprotein	191	100

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
685	BL00266	Somatotropin, prolactin and related hormones proteins.	BL00266A 15.69 8.347e-11 35-61
686	PR00836	SOMATOTROPIN HORMONE FAMILY SIGNATURE	PR00836A 14.40 2.862e-11 79-92 PR00836B 16.59 7.000e-11 101-119
686	BL00266	Somatotropin, prolactin and related hormones proteins.	BL00266B 24.48 8.714e-21 79-116 BL00266A 15.69 1.923e-14 35-61 BL00266D 12.72 4.000e-11 201-224 BL00266C 13.66 3.700e-10 135-151
688	PR00836	SOMATOTROPIN HORMONE FAMILY SIGNATURE	PR00836B 16.59 2.895e-16 101-119 PR00836A 14.40 2.800e-13 79-92
688	BL00266	Somatotropin, prolactin and related hormones proteins.	BL00266B 24.48 4.000e-29 79-116 BL00266A 15.69 9.000e-19 35-61 BL00266D 12.72 4.000e-11 201-224 BL00266C 13.66 4.000e-10 135-151
689	BL00284	Serpins proteins.	BL00284C 28.56 3.700e-26 185-226 BL00284E 19.15 1.333e-17 373-397 BL00284A 15.64 8.714e-16 77-100 BL00284D 16.34 7.279e-12 294-320 BL00284B 17.99 4.825e-10 158-178
690	PR00390	PHOSPHOLIPASE C SIGNATURE	PR00390A 15.09 1.439e-20 191-209
690	BL00303	S-100/ICaBP type calcium binding protein.	BL00303B 26.15 4.971e-09 31-67
690	BL00292	Cyclins proteins.	BL00292A 22.87 5.114e-09 116-149
691	PF00756	Putative esterase.	PF00756C 14.12 1.108e-09 438-467
691	BL00120	Lipases, serine proteins.	BL00120B 11.37 4.462e-09 435-449
693	PR00573	INTERLEUKIN 8B RECEPTOR SIGNATURE	PR00573C 9.99 7.300e-10 38-46
693	PR00427	INTERLEUKIN-8 RECEPTOR SIGNATURE	PR00427A 16.30 9.700e-10 34-48
694	BL01238	GDA1/CD39 family of nucleoside phosphatases proteins.	BL01238A 11.72 8.200e-16 104-118 BL01238D 10.19 4.130e-15 248-261 BL01238C 14.36 6.677e-12 219-240 BL01238B 10.99 2.071e-10 176-186
695	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237F 13.57 5.636e-10 239-263
695	BL00237	G-protein coupled receptors proteins.	BL00237C 13.19 5.034e-12 234-260 BL00237A 27.68 8.600e-10 72-111
695	PR00172	GLUCOSE TRANSPORTER SIGNATURE	PR00172C 9.51 2.612e-09 8-28
696	BL00615	C-type lectin domain proteins.	BL00615A 16.68 2.080e-11 175-192
698	BL01238	GDA1/CD39 family of nucleoside phosphatases proteins.	BL01238A 11.72 4.240e-16 51-65 BL01238D 10.19 2.703e-14 196-209 BL01238C 14.36 2.662e-12 167-188 BL01238B 10.99 6.538e-12 124-134
700	BL00037	Myb DNA-binding domain proteins repeat proteins proteins.	BL00037A 16.68 3.571e-11 231-254
700	PF00569	Zinc finger present in dystrophin, CBP/p300.	PF00569 13.42 4.214e-10 184-200
700	PR00608	CLASS II CYTOCHROME C SIGNATURE	PR00608A 13.74 6.434e-10 118-141
700	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 8.861e-09 123-137 PR00456E 3.06 9.772e-09 122-136
701	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 1.000e-09 280-294
703	PF00650	CRAL/TRIO domain proteins.	PF00650D 24.34 1.776e-12 177-210
703	PR00180	CELLULAR RETINALDEHYDE-BINDING	PR00180A 10.11 7.231e-11 37-59

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		PROTEIN SIGNATURE	PR00180D 12.78 9.769e-10 202-221
705	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 8.286e-09 756-768
705	BL00291	Prion protein.	BL00291A 4.49 8.552e-09 196-230
706	BL00400	LBP / BPI / CETP family proteins.	BL00400D 23.26 7.222e-12 251-287
708	BL00478	LIM domain proteins.	BL00478B 14.79 3.000e-12 31-45
710	BL00604	Synaptophysin / synaptoporin proteins.	BL00604F 5.96 7.718e-10 1379-1423
710	PR00524	CHOLECYSTOKININ TYPE A RECEPTOR SIGNATURE	PR00524F 5.36 7.415e-09 1220-1233
710	BL00242	Integrins alpha chain proteins.	BL00242B 8.13 8.615e-09 469-478
710	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 3.571e-13 1043-1071 BL00420A 20.42 9.082e-13 1125-1153 BL00420A 20.42 2.038e-12 142-170 BL00420A 20.42 4.462e-12 714-742 BL00420A 20.42 8.962e-12 454-482 BL00420A 20.42 9.135e-12 935-963 BL00420A 20.42 9.827e-12 797-825 BL00420A 20.42 1.327e-11 202-230 BL00420A 20.42 3.291e-11 803-831 BL00420A 20.42 3.618e-11 521-549 BL00420A 20.42 4.927e-11 589-617 BL00420A 20.42 6.400e-11 64-92 BL00420A 20.42 8.036e-11 451-479 BL00420A 20.42 8.691e-11 1323-1351 BL00420A 20.42 9.345e-11 199-227 BL00420A 20.42 2.623e-10 944-972 BL00420A 20.42 2.770e-10 100-128 BL00420A 20.42 2.770e-10 842-870 BL00420A 20.42 2.918e-10 741-769 BL00420A 20.42 4.098e-10 1137-1165 BL00420A 20.42 4.393e-10 696-724 BL00420A 20.42 4.541e-10 1170-1198 BL00420A 20.42 5.279e-10 1046-1074 BL00420A 20.42 5.426e-10 296-324 BL00420A 20.42 5.426e-10 1149-1177 BL00420A 20.42 6.754e-10 747-775 BL00420A 20.42 6.754e-10 1061-1089 BL00420A 20.42 6.902e-10 1278-1306 BL00420A 20.42 7.049e-10 624-652 BL00420A 20.42 7.492e-10 1055-1083 BL00420A 20.42 8.082e-10 1037-1065 BL00420A 20.42 8.525e-10 836-864 BL00420A 20.42 8.672e-10 187-215 BL00420A 20.42 8.672e-10 598-626 BL00420A 20.42 8.820e-10 139-167 BL00420A 20.42 8.820e-10 896-924 BL00420A 20.42 8.967e-10 717-745 BL00420A 20.42 9.115e-10 314-342 BL00420A 20.42 9.705e-10 923-951 BL00420A 20.42 9.852e-10 369-397 BL00420A 20.42 9.852e-10 806-834 BL00420A 20.42 9.852e-10 1179-1207

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00420A 20.42 1.138e-09 863-891 BL00420A 20.42 1.415e-09 509-537 BL00420A 20.42 1.415e-09 530-558 BL00420A 20.42 2.523e-09 857-885 BL00420A 20.42 2.800e-09 1182-1210 BL00420A 20.42 2.938e-09 1426-1454 BL00420A 20.42 3.077e-09 630-658 BL00420A 20.42 3.354e-09 103-131 BL00420A 20.42 3.492e-09 782-810 BL00420A 20.42 3.492e-09 1064-1092 BL00420A 20.42 3.631e-09 860-888 BL00420A 20.42 3.769e-09 920-948 BL00420A 20.42 4.185e-09 869-897 BL00420A 20.42 4.600e-09 518-546 BL00420A 20.42 5.015e-09 1317-1345 BL00420A 20.42 5.292e-09 524-552 BL00420A 20.42 5.431e-09 633-661 BL00420A 20.42 5.569e-09 729-757 BL00420A 20.42 5.569e-09 824-852 BL00420A 20.42 5.569e-09 1049-1077 BL00420A 20.42 6.123e-09 366-394 BL00420A 20.42 6.262e-09 491-519 BL00420A 20.42 6.538e-09 914-942 BL00420A 20.42 6.954e-09 566-594 BL00420A 20.42 6.954e-09 711-739 BL00420A 20.42 6.954e-09 893-921 BL00420A 20.42 7.369e-09 818-846 BL00420A 20.42 7.923e-09 1471-1499 BL00420A 20.42 8.062e-09 735-763 BL00420A 20.42 8.477e-09 1347-1375 BL00420A 20.42 8.754e-09 1095-1123 BL00420A 20.42 9.031e-09 61-89 BL00420A 20.42 9.308e-09 311-339 BL00420A 20.42 9.308e-09 938-966 BL00420A 20.42 9.446e-09 1299-1327 BL00420A 20.42 9.585e-09 363-391 BL00420A 20.42 9.723e-09 794-822 BL00420A 20.42 9.862e-09 1302-1330
710	BL01113	C1q domain proteins.	BL01113A 17.99 1.290e-15 423-449 BL01113A 17.99 6.455e-14 1170-1196 BL01113A 17.99 8.909e-14 509-535 BL01113A 17.99 8.909e-14 812-838 BL01113A 17.99 8.909e-14 815-841 BL01113A 17.99 3.676e-13 854-880 BL01113A 17.99 5.622e-13 1040-1066 BL01113A 17.99 8.054e-13 788-814 BL01113A 17.99 9.514e-13 589-615 BL01113A 17.99 9.757e-13 363-389 BL01113A 17.99 1.923e-12 1405-1431 BL01113A 17.99 2.154e-12 845-871 BL01113A 17.99 2.615e-12 932-958 BL01113A 17.99 3.077e-12 953-979

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL01113A 17.99 3.308e-12 524-550 BL01113A 17.99 3.769e-12 566-592 BL01113A 17.99 3.769e-12 797-823 BL01113A 17.99 4.231e-12 624-650 BL01113A 17.99 4.462e-12 1242-1268 BL01113A 17.99 5.154e-12 639-665 BL01113A 17.99 5.846e-12 779-805 BL01113A 17.99 6.308e-12 598-624 BL01113A 17.99 6.538e-12 923-949 BL01113A 17.99 6.538e-12 1046-1072 BL01113A 17.99 7.462e-12 112-138 BL01113A 17.99 7.692e-12 705-731 BL01113A 17.99 8.615e-12 211-237 BL01113A 17.99 8.846e-12 196-222 BL01113A 17.99 9.769e-12 460-486 BL01113A 17.99 1.000e-11 1296-1322 BL01113A 17.99 1.205e-11 1043-1069 BL01113A 17.99 1.409e-11 821-847 BL01113A 17.99 1.614e-11 1182-1208 BL01113A 17.99 1.818e-11 747-773 BL01113A 17.99 3.659e-11 451-477 BL01113A 17.99 4.273e-11 914-940 BL01113A 17.99 4.477e-11 836-862 BL01113A 17.99 4.886e-11 729-755 BL01113A 17.99 5.091e-11 744-770 BL01113A 17.99 5.091e-11 1179-1205 BL01113A 17.99 5.500e-11 633-659 BL01113A 17.99 5.500e-11 714-740 BL01113A 17.99 6.523e-11 1468-1494 BL01113A 17.99 6.727e-11 205-231 BL01113A 17.99 6.727e-11 824-850 BL01113A 17.99 7.341e-11 1423-1449 BL01113A 17.99 8.364e-11 595-621 BL01113A 17.99 9.386e-11 687-713 BL01113A 17.99 9.795e-11 690-716 BL01113A 17.99 1.000e-10 806-832 BL01113A 17.99 1.383e-10 494-520 BL01113A 17.99 1.383e-10 803-829 BL01113A 17.99 1.766e-10 560-586 BL01113A 17.99 1.766e-10 1414-1440 BL01113A 17.99 2.149e-10 938-964 BL01113A 17.99 2.340e-10 208-234 BL01113A 17.99 2.723e-10 64-90 BL01113A 17.99 2.915e-10 372-398 BL01113A 17.99 2.915e-10 592-618 BL01113A 17.99 2.915e-10 1368-1394 BL01113A 17.99 3.298e-10 750-776 BL01113A 17.99 3.872e-10 518-544 BL01113A 17.99 5.404e-10 842-868 BL01113A 17.99 5.596e-10 857-883 BL01113A 17.99 6.170e-10 794-820 BL01113A 17.99 6.745e-10 148-174

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL01113A 17.99 6.745e-10 202-228
			BL01113A 17.99 6.745e-10 1251-1277
			BL01113A 17.99 7.319e-10 929-955
			BL01113A 17.99 7.319e-10 1305-1331
			BL01113A 17.99 7.511e-10 432-458
			BL01113A 17.99 7.702e-10 563-589
			BL01113A 17.99 7.702e-10 896-922
			BL01113A 17.99 8.085e-10 1176-1202
			BL01113A 17.99 8.277e-10 296-322
			BL01113A 17.99 8.660e-10 1317-1343
			BL01113A 17.99 9.234e-10 121-147
			BL01113A 17.99 9.426e-10 863-889
			BL01113A 17.99 1.346e-09 426-452
			BL01113A 17.99 1.519e-09 454-480
			BL01113A 17.99 1.692e-09 500-526
			BL01113A 17.99 1.692e-09 911-937
			BL01113A 17.99 1.865e-09 782-808
			BL01113A 17.99 2.038e-09 1284-1310
			BL01113A 17.99 2.212e-09 94-120
			BL01113A 17.99 2.212e-09 1365-1391
			BL01113A 17.99 2.385e-09 604-630
			BL01113A 17.99 2.385e-09 893-919
			BL01113A 17.99 2.385e-09 1098-1124
			BL01113A 17.99 2.731e-09 1161-1187
			BL01113A 17.99 2.904e-09 1465-1491
			BL01113A 17.99 3.077e-09 506-532
			BL01113A 17.99 3.423e-09 1143-1169
			BL01113A 17.99 3.423e-09 1320-1346
			BL01113A 17.99 3.769e-09 1408-1434
			BL01113A 17.99 3.769e-09 1462-1488
			BL01113A 17.99 3.942e-09 366-392
			BL01113A 17.99 3.942e-09 902-928
			BL01113A 17.99 3.942e-09 1037-1063
			BL01113A 17.99 3.942e-09 1185-1211
			BL01113A 17.99 4.115e-09 1290-1316
			BL01113A 17.99 4.462e-09 557-583
			BL01113A 17.99 4.462e-09 575-601
			BL01113A 17.99 4.981e-09 1055-1081
			BL01113A 17.99 5.154e-09 533-559
			BL01113A 17.99 5.327e-09 678-704
			BL01113A 17.99 5.327e-09 1031-1057
			BL01113A 17.99 5.500e-09 187-213
			BL01113A 17.99 5.500e-09 497-523
			BL01113A 17.99 5.500e-09 1332-1358
			BL01113A 17.99 5.673e-09 329-355
			BL01113A 17.99 5.673e-09 899-925
			BL01113A 17.99 6.192e-09 1006-1032
			BL01113A 17.99 6.192e-09 1155-1181
			BL01113A 17.99 6.365e-09 681-707
			BL01113A 17.99 6.538e-09 723-749
			BL01113A 17.99 6.538e-09 833-859
			BL01113A 17.99 6.712e-09 199-225

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL01113A 17.99 6.712e-09 720-746 BL01113A 17.99 6.885e-09 839-865 BL01113A 17.99 7.058e-09 145-171 BL01113A 17.99 7.058e-09 190-216 BL01113A 17.99 7.231e-09 1236-1262 BL01113A 17.99 7.404e-09 830-856 BL01113A 17.99 7.750e-09 684-710 BL01113A 17.99 7.923e-09 905-931 BL01113A 17.99 8.096e-09 696-722 BL01113A 17.99 8.269e-09 630-656 BL01113A 17.99 8.269e-09 1257-1283 BL01113A 17.99 9.308e-09 299-325 BL01113A 17.99 9.308e-09 944-970 BL01113A 17.99 9.654e-09 457-483 BL01113A 17.99 1.000e-08 67-93 BL01113A 17.99 1.000e-08 908-934
711	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 4.545e-10 211-221
711	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 9.408e-10 3649-3676
711	PR00873	ECHINOIDEA (SEA URCHIN) METALLOTHIONEIN SIGNATURE	PR00873D 8.43 5.500e-09 4326-4344
711	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 4.974e-10 4218-4234 PR00907B 11.29 5.720e-09 162-178
711	BL00425	Arthropod defensins proteins.	BL00425 10.48 5.781e-09 1216-1234
711	PR00261	LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR SIGNATURE	PR00261C 11.37 4.000e-20 1015-1036 PR00261D 12.47 5.125e-20 892-913 PR00261B 14.12 5.588e-20 3600-3621 PR00261B 14.12 9.294e-20 1101-1122 PR00261B 14.12 2.667e-19 1053-1074 PR00261C 11.37 3.250e-19 2852-2873 PR00261A 11.02 7.058e-19 1101-1122 PR00261A 11.02 8.615e-19 1015-1036 PR00261B 14.12 9.500e-19 933-954 PR00261D 12.47 1.500e-18 3721-3742 PR00261B 14.12 2.263e-18 3523-3544 PR00261B 14.12 2.421e-18 2729-2750 PR00261A 11.02 2.833e-18 1144-1165 PR00261D 12.47 3.000e-18 1015-1036 PR00261D 12.47 3.167e-18 1053-1074 PR00261C 11.37 3.618e-18 1053-1074 PR00261A 11.02 5.000e-18 3600-3621 PR00261C 11.37 5.582e-18 2809-2830 PR00261A 11.02 6.000e-18 1053-1074 PR00261C 11.37 6.236e-18 1101-1122 PR00261C 11.37 6.891e-18 3562-3583 PR00261A 11.02 7.000e-18 892-913 PR00261D 12.47 8.167e-18 1144-1165 PR00261D 12.47 8.333e-18 1101-1122 PR00261C 11.37 8.527e-18 3484-3505 PR00261C 11.37 9.018e-18 2767-2788 PR00261C 11.37 1.310e-17 1144-1165 PR00261D 12.47 2.579e-17 3600-3621

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00261B 14.12 2.650e-17 3680-3701 PR00261D 12.47 2.737e-17 3680-3701 PR00261C 11.37 3.017e-17 892-913 PR00261B 14.12 3.250e-17 892-913 PR00261A 11.02 4.158e-17 3562-3583 PR00261F 11.57 5.673e-17 2938-2959 PR00261A 11.02 6.368e-17 2809-2830 PR00261A 11.02 6.684e-17 3680-3701 PR00261A 11.02 6.842e-17 3364-3385 PR00261C 11.37 8.138e-17 3680-3701 PR00261A 11.02 8.895e-17 2729-2750 PR00261C 11.37 9.845e-17 974-995 PR00261D 12.47 1.153e-16 2767-2788 PR00261D 12.47 1.153e-16 3364-3385 PR00261F 11.57 1.321e-16 1015-1036 PR00261D 12.47 1.610e-16 2687-2708 PR00261D 12.47 1.915e-16 974-995 PR00261F 11.57 1.964e-16 2599-2620 PR00261D 12.47 2.831e-16 2852-2873 PR00261B 14.12 2.887e-16 3364-3385 PR00261B 14.12 3.032e-16 2809-2830 PR00261A 11.02 3.136e-16 80-101 PR00261D 12.47 3.441e-16 2809-2830 PR00261D 12.47 3.441e-16 3484-3505 PR00261C 11.37 3.951e-16 2938-2959 PR00261C 11.37 4.246e-16 80-101 PR00261D 12.47 4.356e-16 3523-3544 PR00261E 11.08 5.000e-16 892-913 PR00261C 11.37 5.279e-16 2729-2750 PR00261D 12.47 7.407e-16 80-101 PR00261E 11.08 7.500e-16 3680-3701 PR00261B 14.12 7.532e-16 2767-2788 PR00261A 11.02 7.712e-16 3484-3505 PR00261F 11.57 8.071e-16 1053-1074 PR00261B 14.12 8.403e-16 1015-1036 PR00261C 11.37 8.525e-16 3364-3385 PR00261F 11.57 8.714e-16 3809-3830 PR00261A 11.02 8.932e-16 2767-2788 PR00261F 11.57 9.357e-16 3523-3544 PR00261D 12.47 1.429e-15 2599-2620 PR00261B 14.12 1.554e-15 1144-1165 PR00261A 11.02 1.726e-15 2852-2873 PR00261D 12.47 1.857e-15 933-954 PR00261C 11.37 2.000e-15 3523-3544 PR00261B 14.12 2.108e-15 2599-2620 PR00261B 14.12 2.246e-15 974-995 PR00261F 11.57 2.397e-15 3444-3465 PR00261D 12.47 2.714e-15 3404-3425 PR00261E 11.08 3.211e-15 974-995 PR00261A 11.02 3.323e-15 2687-2708 PR00261E 11.08 3.526e-15 1053-1074 PR00261D 12.47 4.429e-15 3562-3583

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00261E 11.08 4.632e-15 1015-1036 PR00261D 12.47 5.000e-15 2938-2959 PR00261C 11.37 5.286e-15 3404-3425 PR00261E 11.08 5.579e-15 2599-2620 PR00261A 11.02 5.645e-15 3523-3544 PR00261F 11.57 5.966e-15 2638-2659 PR00261B 14.12 6.262e-15 2938-2959 PR00261F 11.57 6.276e-15 2852-2873 PR00261C 11.37 6.286e-15 2638-2659 PR00261E 11.08 6.684e-15 1101-1122 PR00261C 11.37 7.286e-15 3809-3830 PR00261B 14.12 8.062e-15 3444-3465 PR00261E 11.08 8.421e-15 1144-1165 PR00261F 11.57 9.690e-15 2767-2788 PR00261B 14.12 1.000e-14 80-101 PR00261F 11.57 1.145e-14 974-995 PR00261F 11.57 1.581e-14 3364-3385 PR00261A 11.02 2.246e-14 933-954 PR00261C 11.37 2.478e-14 3641-3662 PR00261B 14.12 2.853e-14 3721-3742 PR00261A 11.02 3.631e-14 2938-2959 PR00261D 12.47 3.813e-14 2729-2750 PR00261D 12.47 3.813e-14 3809-3830 PR00261E 11.08 3.850e-14 2767-2788 PR00261E 11.08 4.300e-14 2729-2750 PR00261C 11.37 4.358e-14 3444-3465 PR00261E 11.08 4.450e-14 2938-2959 PR00261D 12.47 4.797e-14 2558-2579 PR00261E 11.08 4.900e-14 3809-3830 PR00261F 11.57 4.919e-14 1101-1122 PR00261F 11.57 5.355e-14 3641-3662 PR00261C 11.37 6.104e-14 2599-2620 PR00261E 11.08 6.400e-14 3641-3662 PR00261A 11.02 7.092e-14 3809-3830 PR00261B 14.12 7.221e-14 3809-3830 PR00261B 14.12 7.353e-14 3641-3662 PR00261F 11.57 7.823e-14 1144-1165 PR00261B 14.12 7.882e-14 2687-2708 PR00261E 11.08 8.350e-14 3721-3742 PR00261E 11.08 8.650e-14 2809-2830 PR00261D 12.47 9.016e-14 3641-3662 PR00261C 11.37 9.328e-14 3721-3742 PR00261D 12.47 9.719e-14 2638-2659 PR00261C 11.37 1.522e-13 3600-3621 PR00261F 11.57 2.688e-13 2729-2750 PR00261E 11.08 2.828e-13 3404-3425 PR00261A 11.02 2.853e-13 2558-2579 PR00261B 14.12 2.901e-13 2852-2873 PR00261E 11.08 2.969e-13 2852-2873 PR00261E 11.08 2.969e-13 3764-3785 PR00261A 11.02 3.515e-13 974-995 PR00261C 11.37 3.609e-13 2687-2708

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00261E 11.08 3.813e-13 3364-3385 PR00261A 11.02 3.912e-13 3721-3742 PR00261E 11.08 4.094e-13 1185-1206 PR00261E 11.08 4.094e-13 2638-2659 PR00261A 11.02 6.162e-13 3404-3425 PR00261A 11.02 6.956e-13 2893-2914 PR00261E 11.08 7.328e-13 3523-3544 PR00261A 11.02 7.485e-13 2599-2620 PR00261F 11.57 7.891e-13 2558-2579 PR00261B 14.12 7.972e-13 2638-2659 PR00261E 11.08 9.016e-13 3562-3583 PR00261E 11.08 9.297e-13 2558-2579 PR00261F 11.57 9.578e-13 3404-3425 PR00261F 11.57 9.578e-13 3680-3701 PR00261D 12.47 1.254e-12 3444-3465 PR00261F 11.57 1.265e-12 2809-2830 PR00261C 11.37 1.370e-12 933-954 PR00261E 11.08 1.545e-12 2687-2708 PR00261F 11.57 1.926e-12 3562-3583 PR00261F 11.57 2.456e-12 3721-3742 PR00261B 14.12 2.603e-12 3562-3583 PR00261F 11.57 3.382e-12 1185-1206 PR00261B 14.12 4.205e-12 3404-3425 PR00261E 11.08 4.955e-12 2893-2914 PR00261A 11.02 5.310e-12 3641-3662 PR00261C 11.37 6.178e-12 125-146 PR00261C 11.37 6.301e-12 1185-1206 PR00261F 11.57 8.147e-12 3484-3505 PR00261E 11.08 8.364e-12 80-101 PR00261E 11.08 8.500e-12 125-146 PR00261B 14.12 8.644e-12 3484-3505 PR00261F 11.57 8.676e-12 892-913 PR00261D 12.47 9.493e-12 2893-2914 PR00261A 11.02 1.365e-11 3444-3465 PR00261F 11.57 1.625e-11 3764-3785 PR00261E 11.08 1.643e-11 3484-3505 PR00261E 11.08 1.771e-11 3600-3621 PR00261A 11.02 2.581e-11 2638-2659 PR00261A 11.02 2.824e-11 1185-1206 PR00261F 11.57 3.500e-11 933-954 PR00261C 11.37 5.263e-11 2558-2579 PR00261F 11.57 5.375e-11 2687-2708 PR00261D 12.47 7.081e-11 125-146 PR00261A 11.02 7.811e-11 125-146 PR00261F 11.57 8.500e-11 3600-3621 PR00261E 11.08 9.871e-11 3444-3465 PR00261F 11.57 2.320e-10 80-101 PR00261F 11.57 2.920e-10 125-146 PR00261C 11.37 3.813e-10 2893-2914 PR00261B 14.12 5.111e-10 2558-2579 PR00261D 12.47 6.377e-10 3764-3785 PR00261D 12.47 6.610e-10 1185-1206

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00261B 14.12 7.667e-10 125-146 PR00261B 14.12 8.889e-10 1185-1206 PR00261A 11.02 8.962e-10 3764-3785 PR00261E 11.08 9.137e-10 933-954 PR00261B 14.12 1.321e-09 2893-2914 PR00261C 11.37 7.429e-09 3764-3785
711	BL01177	Anaphylatoxin domain proteins.	BL01177C 17.39 7.429e-09 2973-2991 BL01177C 17.39 8.286e-09 200-218
711	BL00799	Granulins proteins.	BL00799E 14.64 8.627e-09 1201-1249
711	PR00764	COMPLEMENT C9 SIGNATURE	PR00764B 13.56 3.593e-15 1048-1068 PR00764B 13.56 2.227e-13 3636-3656 PR00764B 13.56 8.091e-13 1139-1159 PR00764B 13.56 5.565e-12 928-948 PR00764B 13.56 7.652e-12 1010-1030 PR00764B 13.56 8.043e-12 3399-3419 PR00764B 13.56 2.250e-11 3595-3615 PR00764B 13.56 4.000e-11 3557-3577 PR00764B 13.56 4.500e-11 2762-2782 PR00764B 13.56 6.000e-11 969-989 PR00764B 13.56 7.125e-11 2633-2653 PR00764B 13.56 8.875e-11 2724-2744 PR00764B 13.56 9.625e-11 887-907 PR00764B 13.56 6.377e-10 2804-2824 PR00764B 13.56 1.338e-09 3479-3499 PR00764B 13.56 1.563e-09 120-140 PR00764B 13.56 3.025e-09 3439-3459 PR00764B 13.56 3.925e-09 75-95 PR00764B 13.56 5.388e-09 2594-2614 PR00764B 13.56 6.963e-09 2553-2573 PR00764B 13.56 8.425e-09 2933-2953 PR00764B 13.56 8.763e-09 3518-3538
711	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 8.412e-15 206-221 BL01187B 12.04 2.333e-12 3019-3034 BL01187B 12.04 7.300e-11 3895-3910 BL01187B 12.04 4.600e-10 2979-2994 BL01187B 12.04 4.825e-09 3855-3870 BL01187A 9.98 5.500e-09 3003-3014 BL01187A 9.98 9.625e-09 190-201
711	BL01209	LDL-receptor class A (LDLRA) domain proteins.	BL01209 9.31 8.313e-16 89-101 BL01209 9.31 9.438e-16 1062-1074 BL01209 9.31 3.368e-15 2818-2830 BL01209 9.31 3.842e-15 1110-1122 BL01209 9.31 4.316e-15 901-913 BL01209 9.31 4.000e-14 2608-2620 BL01209 9.31 4.000e-14 3413-3425 BL01209 9.31 5.125e-14 3571-3583 BL01209 9.31 5.500e-14 1194-1206 BL01209 9.31 7.750e-14 2902-2914 BL01209 9.31 8.125e-14 3650-3662 BL01209 9.31 9.250e-14 1153-1165 BL01209 9.31 1.000e-13 3730-3742 BL01209 9.31 6.700e-13 2738-2750

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL01209 9.31 7.000e-13 3689-3701 BL01209 9.31 8.500e-13 2696-2708 BL01209 9.31 3.605e-12 2567-2579 BL01209 9.31 7.632e-12 3453-3465 BL01209 9.31 8.105e-12 2776-2788 BL01209 9.31 8.579e-12 1024-1036 BL01209 9.31 1.196e-11 2861-2873 BL01209 9.31 3.543e-11 134-146 BL01209 9.31 5.109e-11 3373-3385 BL01209 9.31 6.087e-11 2947-2959 BL01209 9.31 6.478e-11 3609-3621 BL01209 9.31 9.413e-11 3773-3785 BL01209 9.31 1.346e-10 3818-3830 BL01209 9.31 3.769e-10 3493-3505 BL01209 9.31 4.115e-10 3532-3544 BL01209 9.31 4.981e-10 942-954 BL01209 9.31 7.231e-10 983-995 BL01209 9.31 9.679e-09 2647-2659
711	PR00054	FUNGAL ZN-CYS BINUCLEAR CLUSTER SIGNATURE	PR00054B 8.73 1.000e-08 3605-3611
712	BL01209	LDL-receptor class A (LDLRA) domain proteins.	BL01209 9.31 8.313e-16 89-101 BL01209 9.31 3.543e-11 134-146
712	PR00261	LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR SIGNATURE	PR00261A 11.02 3.288e-16 80-101 PR00261C 11.37 9.115e-16 80-101 PR00261D 12.47 3.286e-15 80-101 PR00261B 14.12 5.985e-15 80-101 PR00261C 11.37 6.178e-12 125-146 PR00261E 11.08 8.227e-12 80-101 PR00261E 11.08 8.500e-12 125-146 PR00261F 11.57 6.875e-11 80-101 PR00261D 12.47 7.081e-11 125-146 PR00261A 11.02 7.811e-11 125-146 PR00261F 11.57 2.920e-10 125-146 PR00261B 14.12 7.667e-10 125-146
712	PR00764	COMPLEMENT C9 SIGNATURE	PR00764B 13.56 1.563e-09 120-140
712	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 5.720e-09 162-178
714	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 2.765e-25 233-280 BL00232B 32.79 8.263e-22 458-505 BL00232B 32.79 4.571e-19 1193-1240 BL00232B 32.79 8.857e-19 1083-1130 BL00232B 32.79 2.662e-18 1403-1450 BL00232B 32.79 5.292e-18 979-1026 BL00232B 32.79 9.585e-18 1298-1345 BL00232B 32.79 1.265e-17 672-719 BL00232B 32.79 1.529e-17 118-165 BL00232B 32.79 2.588e-17 776-823 BL00232B 32.79 1.386e-16 876-923 BL00232C 10.65 5.390e-12 1081-1098 BL00232C 10.65 1.391e-11 334-351 BL00232C 10.65 2.174e-11 1296-1313 BL00232C 10.65 4.522e-11 1401-1418 BL00232C 10.65 4.115e-10 977-994

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00232B 32.79 7.200e-10 341-388 BL00232C 10.65 9.827e-10 670-687 BL00232C 10.65 4.474e-09 874-891 BL00232C 10.65 8.737e-09 231-248
714	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 4.353e-11 977-994 PR00205B 11.39 4.529e-11 231-248 PR00205B 11.39 7.529e-11 1081-1098 PR00205B 11.39 1.655e-10 1296-1313 PR00205B 11.39 4.764e-10 1191-1208 PR00205B 11.39 5.091e-10 1401-1418 PR00205B 11.39 6.400e-10 456-473 PR00205B 11.39 1.000e-09 334-351 PR00205B 11.39 1.763e-09 874-891 PR00205B 11.39 7.712e-09 563-580 PR00205B 11.39 9.085e-09 670-687
715	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 2.765e-25 233-280 BL00232B 32.79 8.263e-22 458-505 BL00232B 32.79 4.571e-19 1193-1240 BL00232B 32.79 8.857e-19 1083-1130 BL00232B 32.79 2.662e-18 1403-1450 BL00232B 32.79 5.292e-18 979-1026 BL00232B 32.79 9.585e-18 1298-1345 BL00232B 32.79 1.265e-17 672-719 BL00232B 32.79 1.529e-17 118-165 BL00232B 32.79 2.588e-17 776-823 BL00232B 32.79 1.386e-16 876-923 BL00232C 10.65 5.390e-12 1081-1098 BL00232C 10.65 1.391e-11 334-351 BL00232C 10.65 2.174e-11 1296-1313 BL00232C 10.65 4.522e-11 1401-1418 BL00232C 10.65 4.115e-10 977-994 BL00232B 32.79 7.200e-10 341-388 BL00232C 10.65 9.827e-10 670-687 BL00232C 10.65 4.474e-09 874-891 BL00232C 10.65 8.737e-09 231-248
715	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 4.353e-11 977-994 PR00205B 11.39 4.529e-11 231-248 PR00205B 11.39 7.529e-11 1081-1098 PR00205B 11.39 1.655e-10 1296-1313 PR00205B 11.39 4.764e-10 1191-1208 PR00205B 11.39 5.091e-10 1401-1418 PR00205B 11.39 6.400e-10 456-473 PR00205B 11.39 1.000e-09 334-351 PR00205B 11.39 1.763e-09 874-891 PR00205B 11.39 7.712e-09 563-580 PR00205B 11.39 9.085e-09 670-687
716	BL00708	Prolyl endopeptidase family serine proteins.	BL00708B 24.91 7.197e-12 706-736
716	PF00930	Dipeptidyl peptidase IV (DPP IV) N-terminal region.	PF00930I 15.96 6.373e-17 748-775 PF00930H 20.16 2.482e-13 669-711 PF00930J 8.78 1.000e-11 800-820 PF00930G 21.30 9.613e-09 629-666
717	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 3.118e-14 156-172

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00028 16.07 1.900e-13 352-368 BL00028 16.07 2.565e-12 240-256 BL00028 16.07 4.130e-12 212-228 BL00028 16.07 8.435e-12 324-340 BL00028 16.07 5.154e-11 268-284 BL00028 16.07 6.192e-11 296-312 BL00028 16.07 6.885e-11 184-200
717	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 8.800e-14 172-184 PD00066 13.92 4.857e-12 200-212 PD00066 13.92 5.286e-12 228-240 PD00066 13.92 6.143e-12 340-352 PD00066 13.92 7.000e-12 256-268 PD00066 13.92 2.957e-11 312-324 PD00066 13.92 5.304e-11 50-62 PD00066 13.92 7.231e-10 78-90 PD00066 13.92 3.100e-09 284-296
717	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.909e-15 321-334 PR00048A 10.52 1.000e-14 181-194 PR00048A 10.52 1.000e-14 349-362 PR00048A 10.52 3.571e-13 237-250 PR00048A 10.52 4.857e-13 153-166 PR00048A 10.52 1.947e-11 209-222 PR00048A 10.52 3.842e-11 265-278 PR00048A 10.52 5.737e-11 293-306 PR00048B 6.02 9.308e-11 197-206 PR00048B 6.02 6.063e-10 225-234 PR00048B 6.02 6.063e-10 365-374 PR00048B 6.02 8.875e-10 169-178 PR00048B 6.02 5.737e-09 337-346 PR00048B 6.02 9.053e-09 309-318
718	DM01206	CORONAVIRUS NUCLEOCAPSID PROTEIN.	DM01206B 10.69 3.278e-09 70-89 DM01206B 10.69 4.418e-09 105-124
718	BL00048	Protamine P1 proteins.	BL00048 6.39 7.107e-16 64-90 BL00048 6.39 9.196e-16 63-89 BL00048 6.39 1.132e-12 62-88 BL00048 6.39 2.059e-12 66-92 BL00048 6.39 3.250e-12 65-91 BL00048 6.39 7.618e-12 92-118 BL00048 6.39 2.625e-11 60-86 BL00048 6.39 6.500e-11 113-139 BL00048 6.39 6.750e-11 78-104 BL00048 6.39 6.875e-11 104-130 BL00048 6.39 7.125e-11 112-138 BL00048 6.39 8.625e-11 74-100 BL00048 6.39 2.539e-10 108-134 BL00048 6.39 4.434e-10 61-87 BL00048 6.39 5.855e-10 110-136 BL00048 6.39 6.921e-10 98-124 BL00048 6.39 7.158e-10 109-135 BL00048 6.39 7.750e-10 97-123 BL00048 6.39 8.105e-10 79-105 BL00048 6.39 8.579e-10 19-45 BL00048 6.39 8.934e-10 94-120 BL00048 6.39 9.526e-10 103-129 BL00048 6.39 1.675e-09 101-127 BL00048 6.39

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			1.900e-09 73-99 BL00048 6.39 3.250e-09 81-107 BL00048 6.39 3.475e-09 111-137 BL00048 6.39 3.700e-09 82-108 BL00048 6.39 3.700e-09 96-122 BL00048 6.39 4.263e-09 99-125 BL00048 6.39 5.163e-09 107-133 BL00048 6.39 5.275e-09 67-93 BL00048 6.39 5.275e-09 80-106 BL00048 6.39 5.388e-09 49-75 BL00048 6.39 6.738e-09 116-142 BL00048 6.39 7.975e-09 124-150 BL00048 6.39 8.650e-09 52-78 BL00048 6.39 8.763e-09 18-44 BL00048 6.39 9.100e-09 21-47 BL00048 6.39 9.550e-09 76-102 BL00048 6.39 9.550e-09 100-126 BL00048 6.39 9.663e-09 102-128 BL00048 6.39 1.000e-08 77-103
720	PD01719	PRECURSOR GLYCOPROTEIN SIGNAL RE.	PD01719A 12.89 5.875e-20 1548-1575 PD01719A 12.89 8.200e-17 1719-1746 PD01719A 12.89 9.182e-17 1491-1518 PD01719A 12.89 4.569e-16 1434-1461 PD01719A 12.89 7.286e-14 1605-1632 PD01719A 12.89 2.364e-13 1662-1689
720	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 6.538e-16 2348-2363 BL01187B 12.04 3.647e-15 2191-2206 BL01187B 12.04 5.696e-13 2108-2123 BL01187B 12.04 7.261e-13 2232-2247 BL01187A 9.98 4.316e-11 2172-2183 BL01187A 9.98 1.429e-10 2047-2058 BL01187B 12.04 2.286e-10 2023-2038 BL01187A 9.98 1.750e-09 2332-2343
720	BL01177	Anaphylatoxin domain proteins.	BL01177D 17.50 5.167e-09 2042-2059
720	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 2.256e-10 1000-1023 BL00240B 24.70 5.395e-10 450-473 BL00240B 24.70 3.681e-09 1090-1113 BL00240B 24.70 6.170e-09 634-657
720	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 2.091e-10 2353-2363 PR00010C 11.16 6.357e-09 2196-2206
720	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 6.294e-11 763-795 PD02870B 18.83 8.306e-11 1126-1158 PD02870D 15.74 4.800e-10 1126-1160 PD02870B 18.83 7.400e-10 393-425 PD02870B 18.83 9.600e-10 670-702 PD02870B 18.83 1.862e-09 945-977 PD02870B 18.83 3.585e-09 1215-1247 PD02870D 15.74 6.553e-09 854-888 PD02870B 18.83 6.745e-09 1306-1338
720	BL00281	Bowman-Birk serine protease inhibitors family proteins.	BL00281A 14.18 6.754e-09 2018-2034
720	BL00022	EGF-like domain proteins.	BL00022B 7.54 1.900e-09 2357-2363 BL00022B 7.54 7.300e-09 2200-2206
720	BL00799	Granulins proteins.	BL00799B 11.02 7.429e-09 2014-2049
720	DM00864	EGF-LIKE DOMAIN.	DM00864B 11.34 7.465e-09 2196-2214

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
720	PD02327	GLYCOPROTEIN ANTIGEN PRECURSOR IMMUNOGLO.	PD02327B 19.84 7.818e-09 450-471
720	DM01688	2 POLY-IG RECEPTOR.	DM01688D 13.44 2.756e-09 679-701 DM01688G 16.45 6.040e-09 1210-1241 DM01688D 13.44 8.244e-09 26-48
720	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 5.737e-10 119-128 DM00179 13.97 9.053e-10 494-503 DM00179 13.97 6.870e-09 25-34 DM00179 13.97 8.043e-09 1223-1232 DM00179 13.97 8.435e-09 401-410
720	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 2.479e-11 2344-2360 PR00907B 11.29 3.688e-10 2228-2244 PR00907G 11.63 9.660e-10 2348-2374 PR00907G 11.63 9.745e-10 2232-2258 PR00907G 11.63 9.027e-09 2108-2134
720	PD00015	GLYCOPROTEIN PRECURSOR CELL SL.	PD00015B 5.21 1.000e-08 1279-1285
721	BL00674	AAA-protein family proteins.	BL00674B 4.46 1.122e-09 452-473
721	BL00300	SRP54-type proteins GTP-binding domain proteins.	BL00300B 20.56 3.228e-09 452-497
722	BL00211	ABC transporters family proteins.	BL00211B 13.37 9.053e-22 618-649 BL00211B 13.37 3.314e-13 1430-1461 BL00211A 12.23 2.385e-11 515-526 BL00211A 12.23 1.529e-10 1327-1338
722	PR00326	GTP1/OBG GTP-BINDING PROTEIN FAMILY SIGNATURE	PR00326A 8.75 1.129e-09 513-533 PR00326A 8.75 2.671e-09 1325-1345
722	BL00649	G-protein coupled receptors family 2 proteins.	BL00649F 14.99 4.761e-09 857-878
723	BL00130	Uracil-DNA glycosylase proteins.	BL00130A 13.75 1.000e-08 576-588
724	BL00072	Acyl-CoA dehydrogenases proteins.	BL00072E 24.12 5.014e-12 156-198 BL00072D 30.08 7.136e-10 67-117
725	BL00740	MAM domain proteins.	BL00740A 13.87 7.188e-12 409-421
725	PR00020	MAM DOMAIN SIGNATURE	PR00020A 18.17 9.816e-12 407-425
725	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 4.082e-11 143-159
725	PF00094	von Willebrand factor type D domain proteins.	PF00094A 11.09 5.109e-09 138-147
725	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 7.632e-09 68-93
725	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 9.882e-09 145-171
725	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 9.100e-14 236-251 BL01187B 12.04 5.333e-12 191-206 BL01187B 12.04 6.333e-12 109-124 BL01187A 9.98 9.250e-09 172-183 BL01187A 9.98 1.000e-08 217-228
727	PD00930	PROTEIN GTPASE DOMAIN ACTIVATION.	PD00930B 33.72 6.108e-22 898-938 PD00930A 25.62 3.415e-14 775-800
727	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 4.706e-12 724-739
727	PF00620	GTPase-activator protein for Rho-like GTPases.	PF00620B 14.20 6.000e-10 825-841
727	BL01240	Purine and other phosphorylases family 2 proteins.	BL01240C 25.01 1.414e-09 36-77
729	BL00142	Neutral zinc metallopeptidases, zinc-binding	BL00142 8.38 8.875e-10 412-422

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		region proteins.	
729	PD01719	PRECURSOR GLYCOPROTEIN SIGNAL RE.	PD01719A 12.89 4.150e-15 572-599 PD01719A 12.89 3.487e-10 1222-1249 PD01719A 12.89 6.447e-10 1166-1193 PD01719A 12.89 1.778e-09 1425-1452 PD01719A 12.89 7.556e-09 1091-1118
735	BL00741	Guanine-nucleotide dissociation stimulators CDC24 family sign.	BL00741B 14.27 1.333e-14 302-324
742	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 3.571e-13 656-673 PR00205B 11.39 9.357e-13 233-250 PR00205B 11.39 9.413e-12 339-356 PR00205B 11.39 7.055e-10 450-467 PR00205B 11.39 8.691e-10 553-570
742	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 8.615e-24 235-282 BL00232B 32.79 3.631e-18 555-602 BL00232B 32.79 9.862e-18 452-499 BL00232B 32.79 2.110e-15 125-172 BL00232C 10.65 6.500e-13 233-250 BL00232C 10.65 8.750e-13 656-673 BL00232C 10.65 6.087e-11 339-356 BL00232C 10.65 9.827e-10 450-467
745	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 4.375e-15 216-232 BL00028 16.07 8.313e-15 518-534 BL00028 16.07 1.529e-14 244-260 BL00028 16.07 1.000e-13 188-204 BL00028 16.07 2.350e-13 272-288 BL00028 16.07 1.000e-12 412-428 BL00028 16.07 2.957e-12 356-372 BL00028 16.07 2.957e-12 490-506 BL00028 16.07 2.957e-12 546-562 BL00028 16.07 3.348e-12 384-400 BL00028 16.07 4.522e-12 300-316 BL00028 16.07 6.870e-12 328-344 BL00028 16.07 1.000e-11 160-176 BL00028 16.07 3.400e-10 440-456 BL00028 16.07 1.000e-09 132-148
745	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.091e-15 381-394 PR00048A 10.52 6.727e-15 269-282 PR00048A 10.52 6.727e-15 543-556 PR00048A 10.52 7.545e-15 487-500 PR00048A 10.52 9.182e-15 185-198 PR00048A 10.52 6.143e-13 213-226 PR00048A 10.52 7.429e-13 409-422 PR00048A 10.52 8.714e-13 241-254 PR00048A 10.52 8.714e-13 297-310 PR00048A 10.52 4.706e-12 353-366 PR00048B 6.02 6.000e-12 173-182 PR00048B 6.02 3.077e-11 341-350 PR00048B 6.02 7.923e-11 503-512 PR00048B 6.02 1.000e-10 229-238 PR00048A 10.52 4.522e-10 515-528 PR00048A 10.52 6.870e-10 129-142

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00048B 6.02 8.875e-10 531-540 PR00048A 10.52 1.720e-09 157-170 PR00048A 10.52 2.800e-09 437-450 PR00048B 6.02 2.895e-09 453-462 PR00048B 6.02 5.737e-09 313-322 PR00048A 10.52 6.760e-09 325-338
745	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 5.200e-14 176-188 PD00066 13.92 8.200e-14 344-356 PD00066 13.92 4.000e-13 232-244 PD00066 13.92 1.857e-12 456-468 PD00066 13.92 3.571e-12 534-546 PD00066 13.92 4.000e-12 400-412 PD00066 13.92 1.000e-11 260-272 PD00066 13.92 1.000e-11 372-384 PD00066 13.92 4.522e-11 204-216 PD00066 13.92 1.000e-10 288-300 PD00066 13.92 7.300e-09 506-518
746	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 8.250e-35 37-75
746	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 5.200e-14 251-263 PD00066 13.92 8.200e-14 419-431 PD00066 13.92 4.000e-13 307-319 PD00066 13.92 1.857e-12 531-543 PD00066 13.92 4.000e-12 475-487 PD00066 13.92 1.000e-11 335-347 PD00066 13.92 1.000e-11 447-459 PD00066 13.92 4.522e-11 279-291 PD00066 13.92 1.000e-10 363-375
746	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 4.375e-15 291-307 BL00028 16.07 1.529e-14 319-335 BL00028 16.07 1.000e-13 263-279 BL00028 16.07 2.350e-13 347-363 BL00028 16.07 1.000e-12 487-503 BL00028 16.07 2.957e-12 431-447 BL00028 16.07 3.348e-12 459-475 BL00028 16.07 4.522e-12 375-391 BL00028 16.07 6.870e-12 403-419 BL00028 16.07 1.000e-11 235-251 BL00028 16.07 3.400e-10 515-531 BL00028 16.07 1.000e-09 207-223
746	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.091e-15 456-469 PR00048A 10.52 6.727e-15 344-357 PR00048A 10.52 9.182e-15 260-273 PR00048A 10.52 6.143e-13 288-301 PR00048A 10.52 7.429e-13 484-497 PR00048A 10.52 8.714e-13 316-329 PR00048A 10.52 8.714e-13 372-385 PR00048A 10.52 4.706e-12 428-441 PR00048B 6.02 6.000e-12 248-257 PR00048B 6.02 3.077e-11 416-425 PR00048B 6.02 1.000e-10 304-313 PR00048A 10.52 6.870e-10 204-217

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00048A 10.52 1.720e-09 232-245 PR00048A 10.52 2.800e-09 512-525 PR00048B 6.02 2.895e-09 528-537 PR00048B 6.02 5.737e-09 388-397 PR00048A 10.52 6.760e-09 400-413
747	PF01105	emp24/gp25L/p24 family.	PF01105B 25.12 2.868e-25 144-195
749	PR00405	HIV REV INTERACTING PROTEIN SIGNATURE	PR00405C 19.41 1.000e-18 579-600 PR00405A 17.71 8.147e-18 539-558 PR00405B 11.83 7.300e-17 558-575
749	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 7.688e-09 831-885
751	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 6.143e-21 344-382
751	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 8.500e-13 769-781 PD00066 13.92 4.857e-12 711-723
751	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 4.706e-12 778-791 PR00048B 6.02 6.538e-11 766-775 PR00048A 10.52 1.000e-10 750-763 PR00048A 10.52 4.130e-10 602-615 PR00048B 6.02 6.063e-10 708-717 PR00048A 10.52 8.043e-10 630-643 PR00048A 10.52 8.435e-10 692-705 PR00048A 10.52 1.360e-09 720-733
751	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 3.118e-14 753-769 BL00028 16.07 1.346e-11 781-797 BL00028 16.07 3.769e-11 605-621 BL00028 16.07 9.400e-10 723-739 BL00028 16.07 1.771e-09 695-711
754	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 4.541e-13 790-816
754	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477J 19.04 3.382e-27 1241-1271 BL00477F 17.34 8.500e-25 785-814 BL00477G 19.43 8.826e-23 983-1014 BL00477A 13.50 9.800e-23 122-150 BL00477L 23.51 5.500e-16 1437-1469 BL00477K 17.42 4.529e-14 1382-1405 BL00477E 17.53 6.538e-13 755-775 BL00477B 9.05 6.625e-13 209-221 BL00477I 18.76 2.650e-12 1085-1111 BL00477D 12.73 4.073e-12 729-738 BL00477H 9.07 5.395e-12 1054-1065 BL00477C 15.70 1.161e-10 236-252
755	BL00514	Fibrinogen beta and gamma chains C-terminal domain proteins.	BL00514E 14.28 7.750e-12 299-315 BL00514D 15.35 9.824e-11 280-292 BL00514G 15.98 4.273e-10 362-391 BL00514H 14.95 6.217e-09 397-421
756	BL00790	Receptor tyrosine kinase class V proteins.	BL00790I 20.01 7.638e-10 868-898
756	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 5.309e-09 371-403
756	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 7.261e-09 189-198
756	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014B 14.77 6.400e-10 832-842 PR00014D 12.04 3.700e-09 671-685

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00014C 15.44 4.522e-09 857-875 PR00014D 12.04 8.200e-09 875-889 PR00014D 12.04 9.550e-09 774-788
757	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 2.149e-09 306-329
757	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 1.450e-11 149-162 PR00019B 11.36 5.050e-10 98-111 PR00019B 11.36 7.840e-09 122-135
758	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 2.149e-09 306-329
758	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 1.450e-11 149-162 PR00019B 11.36 5.050e-10 98-111 PR00019B 11.36 7.840e-09 122-135
759	BL00649	G-protein coupled receptors family 2 proteins.	BL00649C 17.82 4.339e-11 1086-1111
759	PR00249	SECRETIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00249C 17.08 4.185e-10 1088-1111
760	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 8.313e-15 277-293 BL00028 16.07 1.900e-13 193-209 BL00028 16.07 6.400e-13 137-153 BL00028 16.07 6.400e-13 389-405 BL00028 16.07 4.913e-12 109-125 BL00028 16.07 8.826e-12 333-349 BL00028 16.07 1.000e-11 361-377 BL00028 16.07 1.692e-11 249-265 BL00028 16.07 3.077e-11 221-237 BL00028 16.07 6.538e-11 305-321 BL00028 16.07 7.577e-11 165-181
760	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 4.000e-14 265-277 PD00066 13.92 5.200e-14 97-109 PD00066 13.92 5.200e-14 293-305 PD00066 13.92 5.200e-14 321-333 PD00066 13.92 2.000e-13 209-221 PD00066 13.92 3.500e-13 181-193 PD00066 13.92 1.000e-12 377-389 PD00066 13.92 4.857e-12 237-249 PD00066 13.92 7.857e-12 125-137 PD00066 13.92 8.826e-11 405-417 PD00066 13.92 5.200e-09 349-361
760	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.500e-14 330-343 PR00048A 10.52 7.000e-14 246-259 PR00048A 10.52 9.250e-14 190-203 PR00048A 10.52 1.643e-13 218-231 PR00048A 10.52 4.857e-13 274-287 PR00048A 10.52 1.000e-12 106-119 PR00048B 6.02 6.000e-12 94-103 PR00048B 6.02 6.000e-12 402-411 PR00048A 10.52 4.789e-11 134-147 PR00048B 6.02 5.846e-11 290-299 PR00048B 6.02 5.846e-11 374-383 PR00048A 10.52 9.526e-11 386-399 PR00048A 10.52 1.391e-10 302-315 PR00048A 10.52 1.783e-10 162-175 PR00048A 10.52 7.261e-10 414-427 PR00048B 6.02 8.875e-10 318-327

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00048B 6.02 5.737e-09 262-271
760	PD02462	PROTEIN BOLA TRANSCRIPTION REGULATION AC.	PD02462A 22.48 1.768e-09 270-304 PD02462A 22.48 6.488e-09 298-332
761	PR00121	SODIUM/POTASSIUM-TRANSPORTING ATPASE SIGNATURE	PR00121D 16.72 6.844e-15 173-194
761	BL00154	E1-E2 ATPases phosphorylation site proteins.	BL00154E 20.37 2.929e-13 446-486 BL00154C 12.38 1.540e-12 176-194
761	PR00119	P-TYPE CATION-TRANSPORTING ATPASE SUPERFAMILY SIGNATURE	PR00119B 13.94 7.245e-12 180-194
761	BL01228	Hypothetical cof family proteins.	BL01228D 17.44 6.348e-09 595-619
763	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 7.686e-09 172-188
764	BL00892	HIT family proteins.	BL00892A 18.17 2.125e-10 177-207
764	BL00064	L-lactate dehydrogenase proteins.	BL00064F 25.14 7.720e-09 295-339
767	PD02102	SUBUNIT E V-ATPASE VACUOLAR ATP SYNTHASE HYDROL.	PD02102A 16.74 8.318e-09 121-164
768	BL00926	Lysyl oxidase copper-binding region proteins.	BL00926E 14.42 2.976e-22 306-342 BL00926D 9.03 6.336e-14 260-306
768	PR00074	LYSYL OXIDASE SIGNATURE	PR00074C 8.72 2.674e-18 311-339 PR00074A 9.55 2.514e-10 255-283
768	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 5.500e-29 33-87 BL00420C 11.90 8.017e-11 118-128 BL00420B 22.67 3.526e-10 147-201
768	PR00258	SPERACT RECEPTOR SIGNATURE	PR00258A 11.46 5.721e-11 139-155 PR00258E 13.33 7.000e-11 117-129 PR00258B 9.63 2.180e-10 48-59 PR00258C 9.05 2.469e-10 63-73 PR00258A 11.46 2.746e-10 29-45 PR00258D 14.41 4.724e-10 94-108 PR00258D 14.41 7.429e-09 210-224
773	BL01315	Phosphatidate cytidylyltransferase proteins.	BL01315C 18.61 1.000e-40 342-385 BL01315A 22.47 8.650e-28 221-252 BL01315B 10.40 1.000e-17 253-266
774	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 5.655e-11 190-204 PR00320C 13.01 8.560e-10 190-204 PR00320B 12.19 8.425e-09 190-204
779	BL01152	Hypothetical hesB/yadR/yfhF family proteins.	BL01152B 20.12 1.581e-17 70-95 BL01152C 25.93 1.659e-11 103-149
783	BL00280	Pancreatic trypsin inhibitor (Kunitz) family proteins.	BL00280 24.61 7.070e-26 547-590
783	PR00453	VON WILLEBRAND FACTOR TYPE A DOMAIN SIGNATURE	PR00453A 12.79 3.483e-14 265-282
783	PR00759	BASIC PROTEASE (KUNITZ-TYPE) INHIBITOR FAMILY SIGNATURE	PR00759C 14.15 1.205e-10 575-590 PR00759B 11.26 7.968e-10 565-575
783	BL01113	C1q domain proteins.	BL01113A 17.99 4.447e-10 54-80 BL01113A 17.99 4.638e-10 100-126 BL01113A 17.99 7.702e-10 57-83 BL01113A 17.99 1.865e-09 106-132 BL01113A 17.99 3.250e-09 60-86 BL01113A 17.99 3.250e-09 213-239 BL01113A 17.99 3.423e-09 34-60 BL01113A 17.99 6.365e-09 198-224

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL01113A 17.99 7.231e-09 109-135
783	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 3.213e-10 16-44 BL00420A 20.42 1.415e-09 100-128 BL00420A 20.42 7.923e-09 216-244 BL00420A 20.42 8.477e-09 169-197
785	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 5.404e-09 336-359
786	PR00918	CALICIVIRUS NON-STRUCTURAL POLYPROTEIN FAMILY SIGNATURE	PR00918A 13.76 4.284e-12 27-47
786	BL01128	Shikimate kinase proteins.	BL01128A 18.84 6.684e-11 394-427
786	BL00795	Involucrin proteins.	BL00795C 17.06 8.000e-11 191-235
786	BL00300	SRP54-type proteins GTP-binding domain proteins.	BL00300B 20.56 4.032e-10 391-436
786	PR00830	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE	PR00830A 8.41 4.452e-09 37-56
786	BL00113	Adenylate kinase proteins.	BL00113A 12.74 3.782e-11 34-50 BL00113B 20.49 4.974e-11 58-101 BL00113A 12.74 5.431e-09 395-411
786	BL00674	AAA-protein family proteins.	BL00674B 4.46 5.986e-09 30-51
786	PR00819	CBXX/CFQX SUPERFAMILY SIGNATURE	PR00819B 10.83 7.247e-09 32-47
786	PR00364	DISEASE RESISTANCE PROTEIN SIGNATURE	PR00364A 8.19 8.057e-09 32-47
786	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 8.914e-09 31-52
788	BL50002	Src homology 3 (SH3) domain proteins profile.	BL50002B 15.18 1.000e-10 42-55 BL50002A 14.19 3.813e-09 4-22
789	BL50002	Src homology 3 (SH3) domain proteins profile.	BL50002B 15.18 1.000e-10 115-128 BL50002A 14.19 3.813e-09 77-95
790	BL00288	Tissue inhibitors of metalloproteinases proteins.	BL00288A 17.47 9.143e-21 10-39 BL00288C 14.62 6.500e-18 73-87 BL00288B 9.44 7.000e-15 54-64
791	BL00615	C-type lectin domain proteins.	BL00615A 16.68 2.080e-11 156-173
792	BL00375	UDP-glycosyltransferases proteins.	BL00375F 16.99 1.000e-40 270-314 BL00375G 13.01 1.000e-40 369-408 BL00375E 18.75 3.250e-37 215-264 BL00375D 14.56 5.622e-24 175-202 BL00375C 18.27 6.478e-24 110-133 BL00375B 21.22 5.000e-22 47-87
794	BL01183	ubiE/COQ5 methyltransferase family proteins.	BL01183B 21.31 6.660e-12 143-187
794	BL01279	Protein-L-isoaspartate(D-aspartate) O-methyltransferase signa.	BL01279A 24.27 5.862e-11 57-104
795	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 3.045e-21 494-533
795	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 2.000e-12 508-530 PR00237B 13.50 4.414e-11 463-484 PR00237D 8.94 5.050e-11 544-565
796	DM01688	2 POLY-IG RECEPTOR.	DM01688B 15.06 2.500e-10 82-129
797	DM01688	2 POLY-IG RECEPTOR.	DM01688B 15.06 2.500e-10 82-129
798	DM01688	2 POLY-IG RECEPTOR.	DM01688B 15.06 3.628e-09 82-129
802	PF00997	Kappa casein.	PF00997D 9.95 8.306e-09 506-540
804	PD02080	T-CELL GLYCOPROTEIN CD8 CHAIN	PD02080B 20.69 9.716e-09 20-58

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		SURFACE ALPHA PRE.	
804	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 9.806e-09 19-58
805	BL00982	Bacterial-type phytoene dehydrogenase proteins.	BL00982E 9.88 4.857e-11 24-39
806	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 8.696e-11 72-97
807	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 8.696e-11 72-97
808	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 8.696e-11 72-97
812	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 2.674e-10 279-302 BL00240B 24.70 8.535e-10 374-397 BL00240B 24.70 7.702e-09 470-493
812	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 4.600e-10 512-544 PD02870B 18.83 7.894e-09 120-152
813	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 4.600e-10 2395-2427 PD02870B 18.83 4.160e-09 1707-1739 PD02870B 18.83 5.883e-09 1806-1838 PD02870B 18.83 7.894e-09 2003-2035 PD02870B 18.83 7.989e-09 435-467
813	PD00015	GLYCOPROTEIN PRECURSOR CELL SI.	PD00015B 5.21 8.000e-09 1481-1487
813	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 2.256e-10 1667-1690 BL00240B 24.70 2.674e-10 2162-2185 BL00240B 24.70 8.535e-10 2257-2280 BL00240B 24.70 4.064e-09 1570-1593 BL00240B 24.70 5.213e-09 300-323 BL00240B 24.70 7.702e-09 2353-2376 BL00240B 24.70 8.851e-09 1473-1496
814	PR00500	POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE	PR00500B 7.74 6.305e-09 220-240
814	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 4.600e-10 2590-2622 PD02870B 18.83 4.160e-09 1902-1934 PD02870B 18.83 5.883e-09 2001-2033 PD02870B 18.83 7.894e-09 2198-2230 PD02870B 18.83 7.989e-09 630-662
814	PD00015	GLYCOPROTEIN PRECURSOR CELL SI.	PD00015B 5.21 8.000e-09 1676-1682
814	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 2.256e-10 1862-1885 BL00240B 24.70 2.674e-10 2357-2380 BL00240B 24.70 8.535e-10 2452-2475 BL00240B 24.70 4.064e-09 1765-1788 BL00240B 24.70 5.213e-09 495-518 BL00240B 24.70 7.702e-09 2548-2571 BL00240B 24.70 8.851e-09 1668-1691
816	PD01733	APOLIPOPROTEIN PLASMA LIPID TRANSPORT H.	PD01733B 20.44 6.600e-14 75-129
816	PD02807	APOLIPOPROTEIN E PRECURSOR APO-E GLYCOPROTEIN PLAS.	PD02807D 7.99 4.779e-09 92-141
817	PD01733	APOLIPOPROTEIN PLASMA LIPID TRANSPORT H.	PD01733B 20.44 6.600e-14 75-129
817	PD02807	APOLIPOPROTEIN E PRECURSOR APO-	PD02807D 7.99 4.779e-09 92-141

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		E GLYCOPROTEIN PLAS.	
819	PR00389	PHOSPHOLIPASE A2 SIGNATURE	PR00389C 18.33 3.172e-20 56-74 PR00389B 10.70 8.154e-15 37-55 PR00389E 12.52 5.385e-14 104-120
819	BL00118	Phospholipase A2 histidine proteins.	BL00118B 16.33 5.875e-33 44-71 BL00118D 12.85 7.500e-14 104-119 BL00118C 13.90 8.342e-10 79-97
821	BL00908	Mandelate racemase / muconate lactonizing enzyme family signa.	BL00908B 37.71 1.900e-15 209-263 BL00908A 15.14 5.310e-10 87-113
822	PF00956	Nuclesosome assembly protein (NAP).	PF00956B 23.14 1.000e-40 99-139 PF00956C 7.72 6.850e-22 153-170 PF00956A 11.88 1.000e-13 58-68 PF00956D 7.51 3.700e-12 232-242
822	BL00824	Elongation factor 1 beta/beta'/delta chain proteins.	BL00824B 9.21 3.676e-09 286-305
823	BL01032	Protein phosphatase 2C proteins.	BL01032C 6.14 3.195e-12 147-156 BL01032H 11.25 5.680e-11 318-330 BL01032G 8.33 8.932e-11 282-295 BL01032I 10.42 8.902e-09 379-388
824	PF00094	von Willebrand factor type D domain proteins.	PF00094C 12.88 1.918e-09 124-133
824	PD02576	PRECURSOR GLYCOPROTEIN SIGNAL CELL.	PD02576A 27.60 9.057e-09 101-149
825	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245C 7.84 5.355e-17 121-136 PR00245B 10.38 3.919e-12 60-74 PR00245E 12.40 1.000e-10 174-188
825	BL00237	G-protein coupled receptors proteins.	BL00237D 11.23 2.091e-09 165-181
825	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 8.714e-11 155-181 PR00237E 13.03 9.735e-09 82-105
826	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245C 7.84 5.355e-17 235-250 PR00245A 18.03 8.615e-15 58-79 PR00245B 10.38 3.919e-12 174-188 PR00245E 12.40 1.000e-10 288-302
826	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 1.581e-15 89-128 BL00237D 11.23 2.091e-09 279-295
826	PR00896	VASOPRESSIN RECEPTOR SIGNATURE	PR00896B 9.01 8.962e-09 54-65
826	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 8.714e-11 269-295 PR00237C 15.69 3.829e-10 103-125 PR00237E 13.03 9.735e-09 196-219
827	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 5.650e-14 39-64 BL00243H 17.53 4.261e-11 5-30
828	PD00930	PROTEIN GTPASE DOMAIN ACTIVATION.	PD00930B 33.72 7.070e-19 201-241
831	PR00193	MYOSIN HEAVY CHAIN SIGNATURE	PR00193C 12.60 1.383e-23 177-204 PR00193B 11.69 2.212e-18 125-150 PR00193A 15.41 5.925e-12 65-84
831	BL00567	Phosphoribulokinase proteins.	BL00567A 10.66 9.031e-10 127-145
832	PR00193	MYOSIN HEAVY CHAIN SIGNATURE	PR00193C 12.60 1.383e-23 177-204 PR00193B 11.69 2.212e-18 125-150 PR00193A 15.41 5.925e-12 65-84
832	BL00567	Phosphoribulokinase proteins.	BL00567A 10.66 9.031e-10 127-145

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
834	BL00484	Thyroglobulin type-1 repeat proteins proteins.	BL00484C 17.01 3.647e-12 358-372 BL00484B 9.04 4.529e-11 338-351
834	BL00282	Kazal serine protease inhibitors family proteins.	BL00282 16.88 3.880e-09 143-165
834	BL00612	Osteonectin domain proteins.	BL00612E 13.12 8.230e-09 274-318
835	BL00817	Erythropoietin / thrombopoietin proteins.	BL00817A 18.03 8.200e-10 515-545
835	PR00251	BACTERIAL OPSIN SIGNATURE	PR00251A 12.15 8.820e-10 515-534
835	PR00807	POLLEN ALLERGEN AMB FAMILY SIGNATURE	PR00807A 16.64 8.151e-09 459-476
836	BL00817	Erythropoietin / thrombopoietin proteins.	BL00817A 18.03 8.200e-10 515-545
836	PR00251	BACTERIAL OPSIN SIGNATURE	PR00251A 12.15 8.820e-10 515-534
836	PR00807	POLLEN ALLERGEN AMB FAMILY SIGNATURE	PR00807A 16.64 8.151e-09 459-476
838	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 8.435e-10 327-340 PR00019A 11.19 9.217e-10 182-195 PR00019A 11.19 3.333e-09 278-291 PR00019B 11.36 3.520e-09 227-240 PR00019B 11.36 9.280e-09 299-312
841	PF00023	Ank repeat proteins.	PF00023A 16.03 6.464e-09 135-150
844	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270D 24.66 5.378e-09 292-327
844	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 9.809e-09 155-178
845	PR00020	MAM DOMAIN SIGNATURE	PR00020A 18.17 5.776e-12 759-777 PR00020C 13.66 6.932e-10 832-843
845	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270D 24.66 5.378e-09 292-327
845	BL00740	MAM domain proteins.	BL00740A 13.87 8.313e-12 761-773 BL00740B 19.76 8.500e-09 901-921
845	PD02080	T-CELL GLYCOPROTEIN CD8 CHAIN SURFACE ALPHA PRE.	PD02080B 20.69 9.621e-09 538-576
845	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 9.809e-09 155-178
847	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 4.273e-09 839-852
847	PF00780	Domain found in NIK1-like kinases, mouse citron and yeast ROM.	PF00780I 14.69 4.825e-09 165-194
848	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 4.273e-09 88-101
851	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 8.250e-12 174-197
851	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 3.842e-10 218-227
851	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 5.500e-10 327-359
851	PR00021	SMALL PROLINE-RICH PROTEIN SIGNATURE	PR00021A 4.31 8.405e-09 402-414
852	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 8.250e-12 170-193
852	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 3.842e-10 214-223
852	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 5.500e-10 323-355
852	PR00021	SMALL PROLINE-RICH PROTEIN SIGNATURE	PR00021A 4.31 8.405e-09 398-410
854	PF00168	C2 domain proteins.	PF00168C 27.49 2.636e-10 183-208 PF00168C 27.49 6.318e-10 316-341
854	PR00399	SYNAPTOTAGMIN SIGNATURE	PR00399C 12.82 7.324e-12 216-231 PR00399A 9.52 8.239e-11 145-160

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00399B 14.27 8.277e-11 160-173 PR00399D 14.48 3.930e-10 236-246 PR00399B 14.27 1.915e-09 291-304
854	PR00360	C2 DOMAIN SIGNATURE	PR00360B 13.61 6.897e-12 200-213 PR00360A 14.59 6.538e-11 304-316 PR00360B 13.61 8.636e-11 333-346 PR00360A 14.59 2.184e-09 173-185
855	PD01719	PRECURSOR GLYCOPROTEIN SIGNAL RE.	PD01719A 12.89 3.483e-16 545-572
855	BL00142	Neutral zinc metallopeptidases, zinc-binding region proteins.	BL00142 8.38 7.545e-11 389-399
855	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 9.182e-10 384-402
857	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 3.077e-09 58-72
857	PF00930	Dipeptidyl peptidase IV (DPP IV) N-terminal region.	PF00930J 8.78 1.000e-08 267-287
858	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 3.077e-09 51-65
858	PF00930	Dipeptidyl peptidase IV (DPP IV) N-terminal region.	PF00930J 8.78 1.000e-08 260-280
859	PR00258	SPERACT RECEPTOR SIGNATURE	PR00258A 11.46 8.054e-16 333-349 PR00258B 9.63 1.509e-12 352-363 PR00258E 13.33 1.833e-10 421-433
859	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 7.582e-30 337-391 BL00420C 11.90 9.100e-13 422-432 BL00420A 20.42 8.269e-12 249-277 BL00420A 20.42 7.382e-11 264-292 BL00420A 20.42 1.885e-10 288-316 BL00420A 20.42 7.344e-10 246-274 BL00420A 20.42 2.246e-09 261-289
859	BL01113	C1q domain proteins.	BL01113A 17.99 3.189e-13 264-290 BL01113A 17.99 5.909e-11 246-272 BL01113A 17.99 1.383e-10 273-299 BL01113A 17.99 2.149e-10 258-284 BL01113A 17.99 2.915e-10 261-287 BL01113A 17.99 5.596e-10 252-278 BL01113A 17.99 7.128e-10 267-293 BL01113A 17.99 1.692e-09 282-308 BL01113A 17.99 5.154e-09 255-281
860	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 8.333e-39 397-451 BL00420C 11.90 9.100e-13 482-492 BL00420A 20.42 9.135e-12 309-337 BL00420A 20.42 7.382e-11 324-352 BL00420A 20.42 1.885e-10 348-376 BL00420A 20.42 7.639e-10 306-334 BL00420A 20.42 2.246e-09 321-349
860	PR00258	SPERACT RECEPTOR SIGNATURE	PR00258A 11.46 8.054e-16 393-409 PR00258B 9.63 1.509e-12 412-423 PR00258E 13.33 1.833e-10 481-493 PR00258C 9.05 3.667e-09 427-437
860	BL01113	C1q domain proteins.	BL01113A 17.99 3.189e-13 324-350

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL01113A 17.99 5.295e-11 306-332 BL01113A 17.99 1.383e-10 333-359 BL01113A 17.99 2.149e-10 318-344 BL01113A 17.99 2.915e-10 321-347 BL01113A 17.99 7.128e-10 327-353 BL01113A 17.99 1.692e-09 342-368 BL01113A 17.99 4.115e-09 312-338 BL01113A 17.99 5.673e-09 315-341
862	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 1.450e-13 222-238 BL00028 16.07 1.000e-12 474-490 BL00028 16.07 8.435e-12 502-518 BL00028 16.07 1.346e-11 306-322 BL00028 16.07 2.731e-11 362-378 BL00028 16.07 2.731e-11 390-406 BL00028 16.07 3.423e-11 250-266 BL00028 16.07 3.423e-11 334-350 BL00028 16.07 7.577e-11 418-434 BL00028 16.07 1.600e-10 194-210 BL00028 16.07 9.400e-10 278-294
862	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 8.200e-16 322-334 PD00066 13.92 7.231e-15 406-418 PD00066 13.92 7.923e-15 462-474 PD00066 13.92 4.600e-14 378-390 PD00066 13.92 5.200e-14 490-502 PD00066 13.92 1.000e-13 210-222 PD00066 13.92 1.000e-13 294-306 PD00066 13.92 3.000e-13 238-250 PD00066 13.92 5.304e-11 266-278 PD00066 13.92 7.652e-11 350-362 PD00066 13.92 7.000e-09 434-446
862	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 7.545e-15 415-428 PR00048A 10.52 2.929e-13 387-400 PR00048A 10.52 6.786e-13 219-232 PR00048A 10.52 8.714e-13 443-456 PR00048A 10.52 2.059e-12 247-260 PR00048A 10.52 2.059e-12 331-344 PR00048A 10.52 5.235e-12 471-484 PR00048A 10.52 9.471e-12 499-512 PR00048B 6.02 2.385e-11 319-328 PR00048B 6.02 2.385e-11 487-496 PR00048A 10.52 9.053e-11 303-316 PR00048B 6.02 1.563e-10 375-384 PR00048A 10.52 2.957e-10 359-372 PR00048A 10.52 3.348e-10 191-204 PR00048B 6.02 8.313e-10 459-468 PR00048A 10.52 9.217e-10 275-288 PR00048B 6.02 9.438e-10 207-216 PR00048B 6.02 1.947e-09 263-272 PR00048B 6.02 3.368e-09 235-244 PR00048B 6.02 3.368e-09 291-300 PR00048B 6.02 7.158e-09 403-412
863	PD01234	PROTEIN NUCLEAR BROMODOMAIN	PD01234B 15.53 3.250e-09 568-585

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		TRANS.	
865	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320B 12.19 1.257e-10 225-239 PR00320A 16.74 4.441e-10 225-239
865	BL00678	Trp-Asp (WD) repeat proteins proteins.	BL00678 9.67 9.053e-09 227-237
867	BL00600	Aminotransferases class-III pyridoxal-phosphate attachment si.	BL00600E 16.43 1.771e-17 302-330 BL00600A 17.98 3.880e-17 98-121 BL00600G 12.43 9.625e-17 377-395 BL00600B 19.60 5.091e-15 160-185 BL00600F 8.77 2.421e-12 343-355 BL00600C 16.18 6.040e-12 190-205 BL00600D 8.71 1.000e-10 281-294
868	BL00600	Aminotransferases class-III pyridoxal-phosphate attachment si.	BL00600E 16.43 1.771e-17 199-227 BL00600G 12.43 9.625e-17 274-292 BL00600B 19.60 2.703e-14 57-82 BL00600F 8.77 2.421e-12 240-252 BL00600C 16.18 6.040e-12 87-102 BL00600D 8.71 1.000e-10 178-191
869	BL00021	Kringle domain proteins.	BL00021D 24.56 1.188e-24 248-289 BL00021B 13.33 2.983e-13 88-105
869	BL00134	Serine proteases, trypsin family, histidine proteins.	BL00134C 13.45 8.800e-15 276-289 BL00134A 11.96 9.438e-15 88-104 BL00134B 15.99 3.676e-12 237-260
869	BL00495	Apple domain proteins.	BL00495O 13.75 8.597e-16 267-295 BL00495N 11.04 2.235e-11 229-263 BL00495K 12.58 4.990e-10 90-122
869	PR00722	CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1) SIGNATURE	PR00722C 10.87 3.571e-14 236-248 PR00722A 12.27 5.966e-14 89-104 PR00722B 12.51 9.571e-10 145-159
869	BL01253	Type I fibronectin domain proteins.	BL01253H 13.15 3.609e-23 258-292 BL01253G 11.34 4.103e-15 236-249 BL01253D 4.84 4.360e-09 88-101
870	BL00188	Biotin-requiring enzymes attachment site proteins.	BL00188 30.29 9.122e-09 154-199
873	DM00758	AGRIN.	DM00758 13.12 6.459e-10 93-108
873	BL00612	Osteonectin domain proteins.	BL00612B 11.35 1.284e-09 86-118
873	DM00060	338 kw NEUREXIN ALPHA III CYSTEINE.	DM00060 6.92 8.000e-11 1048-1057 DM00060 6.92 4.060e-09 128-137
873	BL01185	C-terminal cystine knot proteins.	BL01185B 21.14 4.388e-09 234-282
873	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010A 11.79 1.450e-12 46-57 PR00010C 11.16 2.333e-11 184-194 PR00010C 11.16 9.333e-11 296-306 PR00010C 11.16 4.273e-10 66-76 PR00010C 11.16 7.000e-10 28-38 PR00010A 11.79 7.097e-10 488-499 PR00010C 11.16 3.571e-09 546-556 PR00010A 11.79 4.231e-09 564-575 PR00010C 11.16 5.929e-09 374-384
873	PR00764	COMPLEMENT C9 SIGNATURE	PR00764F 16.89 4.699e-10 52-72 PR00764F 16.89 5.562e-10 170-190 PR00764F 16.89 6.301e-10 321-341 PR00764F 16.89 9.753e-10 360-380

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00764F 16.89 2.052e-09 570-590 PR00764F 16.89 2.636e-09 398-418 PR00764F 16.89 7.312e-09 128-148 PR00764F 16.89 7.662e-09 282-302 PR00764F 16.89 7.662e-09 532-552
873	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011B 13.08 6.425e-09 63-81 PR00011B 13.08 8.521e-09 25-43
873	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 8.531e-09 75-120
873	BL00022	EGF-like domain proteins.	BL00022B 7.54 1.000e-09 378-384 BL00022A 7.48 9.000e-09 173-179 BL00022A 7.48 9.000e-09 363-369
873	BL00279	Membrane attack complex components / perforin proteins.	BL00279E 37.11 2.000e-13 553-600 BL00279E 37.11 6.875e-13 343-390 BL00279E 37.11 6.803e-12 1031-1078 BL00279E 37.11 2.962e-11 35-82 BL00279E 37.11 5.731e-11 304-351 BL00279E 37.11 7.115e-11 73-120 BL00279E 37.11 7.462e-11 515-562 BL00279E 37.11 1.217e-10 265-312 BL00279E 37.11 4.349e-09 153-200 BL00279E 37.11 9.163e-09 381-428
873	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 3.333e-12 541-556 BL01187B 12.04 4.000e-12 179-194 BL01187B 12.04 8.000e-12 291-306 BL01187B 12.04 4.300e-11 617-632 BL01187B 12.04 7.900e-11 407-422 BL01187B 12.04 1.514e-10 23-38 BL01187B 12.04 3.829e-10 369-384 BL01187B 12.04 5.371e-10 503-518 BL01187B 12.04 7.171e-10 137-152 BL01187A 9.98 7.429e-10 486-497 BL01187B 12.04 7.429e-10 61-76 BL01187B 12.04 2.800e-09 1057-1072 BL01187B 12.04 3.475e-09 579-594 BL01187A 9.98 4.375e-09 44-55 BL01187B 12.04 7.300e-09 255-270 BL01187B 12.04 9.550e-09 330-345
873	PD00919	CALCIUM-BINDING PRECURSOR SIGNAL R.	PD00919A 11.53 8.820e-10 280-291 PD00919A 11.53 9.864e-09 568-579
874	PR00960	LMBP PROTEIN SIGNATURE	PR00960A 10.63 4.667e-09 78-93
875	BL00738	S-adenosyl-L-homocysteine hydrolase proteins.	BL00738J 18.61 1.000e-40 459-508 BL00738H 23.08 5.320e-36 335-387 BL00738F 12.23 7.261e-29 254-285 BL00738A 16.27 9.660e-27 83-122 BL00738C 16.53 7.923e-25 148-185 BL00738G 14.29 6.268e-23 313-334 BL00738B 12.28 8.085e-21 123-147 BL00738E 14.18 9.200e-19 228-250 BL00738I 14.57 5.135e-17 412-449 BL00738D 7.16 5.109e-13 202-216
875	BL00836	Alanine dehydrogenase & pyridine nucleotide transhydrogenase.	BL00836D 22.30 8.622e-09 291-327

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
877	PR00425	BRADYKININ RECEPTOR SIGNATURE	PR00425C 13.23 3.586e-09 426-445
878	BL00514	Fibrinogen beta and gamma chains C-terminal domain proteins.	BL00514C 17.41 2.579e-24 181-217 BL00514G 15.98 9.111e-12 324-353 BL00514F 11.65 8.914e-09 271-285 BL00514D 15.35 9.565e-09 222-234
879	BL00514	Fibrinogen beta and gamma chains C-terminal domain proteins.	BL00514C 17.41 2.579e-24 181-217 BL00514G 15.98 9.111e-12 324-353 BL00514F 11.65 8.914e-09 271-285 BL00514D 15.35 9.565e-09 222-234
880	BL00514	Fibrinogen beta and gamma chains C-terminal domain proteins.	BL00514C 17.41 2.579e-24 181-217 BL00514G 15.98 9.111e-12 324-353 BL00514F 11.65 8.914e-09 271-285 BL00514D 15.35 9.565e-09 222-234
883	BL00218	Amino acid permeases proteins.	BL00218D 21.49 7.446e-11 244-288 BL00218E 23.30 3.640e-10 325-364
884	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 3.172e-11 158-188
885	BL00615	C-type lectin domain proteins.	BL00615A 16.68 6.538e-10 41-58
889	BL00216	Sugar transport proteins.	BL00216B 27.64 4.900e-10 239-288
891	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 9.526e-10 118-127
891	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 1.305e-09 155-169 PR00049D 0.00 6.797e-09 156-170
892	BL00633	Bromodomain proteins.	BL00633B 13.82 5.950e-21 95-119 BL00633A 14.69 5.154e-14 74-86 BL00633C 15.24 8.071e-14 421-433 BL00633B 13.82 4.600e-13 388-412
892	DM00406	GLIADIN.	DM00406 7.73 5.135e-10 970-982 DM00406 7.73 8.054e-10 753-765
892	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.866e-11 755-769 PR00049D 0.00 9.471e-11 756-770 PR00049D 0.00 2.220e-09 748-762 PR00049D 0.00 3.288e-09 972-986
892	DM00250	kw ANNEXIN ANTIGEN PROLINE TUMOR.	DM00250B 13.84 8.031e-11 1009-1032 DM00250A 10.52 6.607e-09 772-787 DM00250B 13.84 7.568e-09 754-777 DM00250B 13.84 7.689e-09 755-778
892	PR00021	SMALL PROLINE-RICH PROTEIN SIGNATURE	PR00021A 4.31 3.734e-09 967-979 PR00021A 4.31 6.582e-09 771-783 PR00021A 4.31 7.722e-09 769-781
892	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 7.750e-09 255-267
892	BL00415	Synapsins proteins.	BL00415N 4.29 3.231e-12 749-792 BL00415N 4.29 6.504e-12 750-793 BL00415N 4.29 4.857e-11 748-791 BL00415N 4.29 1.824e-10 1003-1046 BL00415N 4.29 6.221e-10 1002-1045 BL00415N 4.29 9.313e-10 964-1007 BL00415N 4.29 2.314e-09 958-1001 BL00415P 2.37 8.200e-09 747-782
892	PR00209	ALPHA/BETA GLIADIN FAMILY SIGNATURE	PR00209B 4.88 3.837e-10 966-984 PR00209B 4.88 5.696e-10 968-986 PR00209B 4.88 8.141e-10 752-770

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00209B 4.88 8.594e-09 758-776
892	BL00904	Protein prenyltransferases alpha subunit repeat proteins proteins.	BL00904A 8.30 5.340e-09 768-817 BL00904A 8.30 9.489e-09 752-801
892	PD02059	CORE POLYPROTEIN PROTEIN GAG CONTAINS: P.	PD02059B 24.48 9.746e-09 867-901
892	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 2.313e-12 750-782 DM00215 19.43 7.000e-12 748-780 DM00215 19.43 9.438e-12 754-786 DM00215 19.43 7.000e-11 749-781 DM00215 19.43 8.412e-11 752-784 DM00215 19.43 1.161e-10 953-985 DM00215 19.43 7.429e-10 948-980 DM00215 19.43 1.000e-09 751-783 DM00215 19.43 2.678e-09 759-791 DM00215 19.43 3.441e-09 753-785 DM00215 19.43 4.508e-09 240-272 DM00215 19.43 4.661e-09 241-273 DM00215 19.43 4.966e-09 765-797 DM00215 19.43 6.492e-09 954-986 DM00215 19.43 8.322e-09 945-977 DM00215 19.43 9.847e-09 747-779
892	PR00503	BROMODOMAIN SIGNATURE	PR00503D 20.81 1.409e-18 421-440 PR00503B 9.96 7.750e-18 94-110 PR00503C 19.84 1.720e-15 110-128 PR00503A 14.39 6.824e-13 78-91 PR00503B 9.96 4.400e-12 387-403 PR00503D 20.81 1.188e-11 128-147 PR00503C 19.84 1.000e-08 403-421
894	BL00282	Kazal serine protease inhibitors family proteins.	BL00282 16.88 2.397e-14 92-114
894	PR00290	KAZAL-TYPE SERINE PROTEASE INHIBITOR SIGNATURE	PR00290A 10.88 2.286e-11 92-102
894	PR00450	RECOVERIN FAMILY SIGNATURE	PR00450C 12.22 4.532e-09 182-203
895	PR00753	1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE SIGNATURE	PR00753E 8.01 8.522e-11 171-195
896	BL00478	LIM domain proteins.	BL00478B 14.79 4.000e-12 102-116 BL00478B 14.79 6.000e-12 173-187 BL00478B 14.79 6.200e-11 43-57 BL00478B 14.79 9.135e-10 231-245
897	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 5.787e-13 467-485
897	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479C 12.01 7.300e-13 512-524
897	BL00239	Receptor tyrosine kinase class II proteins.	BL00239B 25.15 8.948e-13 402-449
897	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 9.217e-14 467-497 BL00107B 13.31 8.714e-11 533-548
897	PF00564	Octicosapeptide repeat proteins.	PF00564B 24.74 6.442e-09 418-468
898	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 5.787e-13 654-672
898	BL00479	Phorbol esters / diacylglycerol binding	BL00479C 12.01 7.300e-13 699-711

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		domain proteins.	
898	BL00239	Receptor tyrosine kinase class II proteins.	BL00239B 25.15 8.948e-13 589-636
898	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 9.217e-14 654-684 BL00107B 13.31 8.714e-11 720-735
898	PF00564	Octicosapeptide repeat proteins.	PF00564B 24.74 6.442e-09 605-655
900	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007C 15.60 3.893e-18 199-220 PR00007A 19.33 7.500e-17 124-150 PR00007B 14.16 2.688e-16 151-170 PR00007D 9.64 5.154e-11 232-242
900	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 6.400e-11 77-105 BL00420A 20.42 6.164e-10 25-53 BL00420A 20.42 9.262e-10 68-96 BL00420A 20.42 1.277e-09 65-93
900	BL01113	C1q domain proteins.	BL01113B 18.26 8.031e-28 130-165 BL01113C 13.18 7.000e-18 199-218 BL01113A 17.99 5.135e-13 95-121 BL01113D 7.47 7.231e-12 234-243 BL01113A 17.99 3.864e-11 34-60 BL01113A 17.99 1.191e-10 71-97 BL01113A 17.99 1.957e-10 77-103 BL01113A 17.99 1.000e-09 28-54 BL01113A 17.99 5.154e-09 68-94 BL01113A 17.99 7.577e-09 74-100 BL01113A 17.99 8.615e-09 83-109
901	PR00927	ADENINE NUCLEOTIDE TRANSLOCATOR 1 SIGNATURE	PR00927A 7.98 9.667e-09 14-26
902	PR00209	ALPHA/BETA GLIADIN FAMILY SIGNATURE	PR00209B 4.88 4.494e-12 427-445
902	BL00415	Synapsins proteins.	BL00415N 4.29 6.771e-10 425-468
902	PR00021	SMALL PROLINE-RICH PROTEIN SIGNATURE	PR00021A 4.31 3.278e-09 448-460
902	DM00406	GLIADIN.	DM00406 7.73 3.919e-10 427-439 DM00406 7.73 6.400e-09 448-460
902	PR00208	GLIADIN AND LMW GLUTENIN SUPERFAMILY SIGNATURE	PR00208A 12.59 5.438e-09 402-419 PR00208A 12.59 7.534e-09 420-437 PR00208A 12.59 8.521e-09 419-436
902	BL00795	Involucrin proteins.	BL00795C 17.06 1.105e-10 396-440 BL00795C 17.06 6.651e-10 411-455 BL00795C 17.06 6.965e-10 394-438 BL00795C 17.06 7.698e-10 422-466 BL00795C 17.06 2.900e-09 408-452 BL00795C 17.06 3.800e-09 395-439 BL00795C 17.06 5.200e-09 425-469 BL00795C 17.06 9.200e-09 424-468
905	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 8.435e-10 5-18
908	BL01208	VWFC domain proteins.	BL01208B 15.83 3.250e-10 1480-1494
908	PR00457	ANIMAL HAEM PEROXIDASE SIGNATURE	PR00457E 20.67 3.118e-22 1041-1067 PR00457D 16.81 4.194e-21 1016-1036 PR00457C 19.25 1.675e-13 998-1016 PR00457H 15.90 5.680e-13 1292-1306 PR00457F 13.69 4.750e-12 1094-1104

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00457G 17.45 8.615e-12 1221-1241 PR00457B 13.29 3.411e-10 846-861
908	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 1.000e-09 325-348
908	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 4.581e-09 304-343
908	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 7.480e-09 73-86
909	BL01208	VWFC domain proteins.	BL01208B 15.83 3.250e-10 1511-1525
909	PR00457	ANIMAL HAEM PEROXIDASE SIGNATURE	PR00457E 20.67 3.118e-22 1072-1098 PR00457D 16.81 4.194e-21 1047-1067 PR00457C 19.25 1.675e-13 1029-1047 PR00457H 15.90 5.680e-13 1323-1337 PR00457F 13.69 4.750e-12 1125-1135 PR00457G 17.45 8.615e-12 1252-1272 PR00457B 13.29 3.411e-10 877-892
909	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 1.000e-09 356-379
909	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 4.581e-09 335-374
909	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 7.480e-09 104-117
910	BL01208	VWFC domain proteins.	BL01208B 15.83 3.250e-10 1373-1387
910	PR00457	ANIMAL HAEM PEROXIDASE SIGNATURE	PR00457E 20.67 3.118e-22 934-960 PR00457D 16.81 4.194e-21 909-929 PR00457C 19.25 1.675e-13 891-909 PR00457H 15.90 5.680e-13 1185-1199 PR00457F 13.69 4.750e-12 987-997 PR00457G 17.45 8.615e-12 1114-1134 PR00457B 13.29 3.411e-10 739-754
910	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 1.000e-09 302-325
910	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 7.677e-09 281-320
910	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 8.920e-09 73-86
911	BL00022	EGF-like domain proteins.	BL00022B 7.54 3.250e-10 881-887 BL00022B 7.54 1.000e-09 88-94
911	PR00764	COMPLEMENT C9 SIGNATURE	PR00764F 16.89 8.274e-10 942-962 PR00764F 16.89 6.377e-09 576-596
911	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010A 11.79 3.700e-12 43-54 PR00010C 11.16 5.636e-10 84-94 PR00010C 11.16 6.727e-10 122-132 PR00010A 11.79 8.258e-10 168-179 PR00010A 11.79 1.231e-09 102-113 PR00010C 11.16 5.500e-09 877-887 PR00010C 11.16 7.000e-09 230-240
911	DM00060	338 kw NEUREXIN ALPHA III CYSTEINE.	DM00060 6.92 7.250e-11 942-951 DM00060 6.92 8.740e-09 576-585
911	BL00279	Membrane attack complex components / perforin proteins.	BL00279E 37.11 1.000e-10 925-972 BL00279E 37.11 4.470e-10 846-893 BL00279E 37.11 8.744e-09 559-606
911	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 9.667e-12 117-132 BL01187A 9.98 9.053e-11 166-177 BL01187B 12.04 6.175e-09 834-849 BL01187A 9.98 8.125e-09 41-52 BL01187B 12.04 9.325e-09 183-198

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
911	PD00919	CALCIUM-BINDING PRECURSOR SIGNAL R.	PD00919A 11.53 9.410e-10 574-585 PD00919A 11.53 9.864e-09 47-58
914	BL00888	Cyclic nucleotide-binding domain proteins.	BL00888B 14.79 4.000e-16 161-184 BL00888B 14.79 1.692e-14 279-302
914	DM01513	CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN.	DM01513B 6.81 8.457e-34 198-249 DM01513B 6.81 2.500e-14 322-373
914	PR00103	CAMP-DEPENDENT PROTEIN KINASE SIGNATURE	PR00103B 13.39 1.000e-16 173-187 PR00103A 9.59 8.105e-15 276-290 PR00103E 17.80 9.591e-15 355-367 PR00103D 10.83 3.700e-14 334-345 PR00103B 13.39 5.935e-13 291-305 PR00103A 9.59 1.500e-12 158-172 PR00103C 15.68 1.000e-11 322-331 PR00103D 10.83 4.349e-10 210-221
915	PD00289	PROTEIN SH3 DOMAIN REPEAT PRESYNA.	PD00289 9.97 8.920e-10 602-615
916	PR00087	LIPOXYGENASE SIGNATURE	PR00087C 15.00 3.057e-21 373-393 PR00087A 18.37 7.955e-18 335-352 PR00087B 15.25 1.000e-16 353-370
916	BL00711	Lipoxygenases iron-binding region proteins.	BL00711E 19.66 8.909e-35 364-400 BL00711I 18.56 4.250e-34 526-563 BL00711D 17.56 2.800e-24 296-321 BL00711H 23.34 5.091e-23 484-522 BL00711C 20.75 2.227e-21 221-249 BL00711F 19.79 5.065e-16 434-450 BL00711B 14.24 1.290e-15 160-175 BL00711G 21.83 8.636e-12 452-483 BL00711A 15.87 5.645e-11 94-103
916	PR00467	MAMMALIAN LIPOXYGENASE SIGNATURE	PR00467F 11.25 4.661e-18 418-440 PR00467E 9.00 5.500e-17 293-312 PR00467A 8.04 4.000e-13 11-28 PR00467D 16.69 5.210e-12 196-217 PR00467B 17.25 1.831e-11 57-76 PR00467C 12.06 1.662e-09 134-148
917	PR00467	MAMMALIAN LIPOXYGENASE SIGNATURE	PR00467E 9.00 5.500e-17 266-285 PR00467A 8.04 4.000e-13 11-28 PR00467D 16.69 5.210e-12 169-190 PR00467B 17.25 1.831e-11 57-76
917	BL00711	Lipoxygenases iron-binding region proteins.	BL00711C 20.75 2.227e-21 194-222 BL00711B 14.24 1.290e-15 131-146 BL00711A 15.87 5.645e-11 94-103
918	BL00711	Lipoxygenases iron-binding region proteins.	BL00711C 20.75 2.227e-21 223-251 BL00711B 14.24 1.290e-15 160-175 BL00711A 15.87 5.645e-11 94-103
918	PR00467	MAMMALIAN LIPOXYGENASE SIGNATURE	PR00467E 9.00 5.500e-17 295-314 PR00467A 8.04 4.000e-13 11-28 PR00467D 16.69 5.210e-12 198-219 PR00467B 17.25 1.831e-11 57-76 PR00467C 12.06 1.662e-09 134-148
927	PD00919	CALCIUM-BINDING PRECURSOR SIGNAL R.	PD00919A 11.53 8.377e-10 216-227

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
927	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 7.429e-10 108-123 BL01187B 12.04 9.486e-10 189-204 BL01187B 12.04 2.800e-09 227-242
927	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011D 14.03 4.158e-12 39-57 PR00011B 13.08 2.973e-09 39-57
927	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 7.276e-09 65-90
927	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 5.929e-09 194-204 PR00010C 11.16 8.286e-09 113-123
927	BL01185	C-terminal cystine knot proteins.	BL01185B 21.14 9.047e-09 168-216
927	DM00060	338 kw NEUREXIN ALPHA III CYSTEINE.	DM00060 6.92 9.460e-09 139-148
927	BL01248	Laminin-type EGF-like (LE) domain proteins.	BL01248 11.02 9.660e-09 48-60
928	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 7.835e-09 1-15
933	BL00680	Methionine aminopeptidase subfamily 1 proteins.	BL00680 14.37 5.304e-17 173-194
933	BL01202	Methionine aminopeptidase subfamily 2 proteins.	BL01202B 26.24 9.671e-10 173-210
933	PR00599	METHIONINE AMINOPEPTIDASE-1 SIGNATURE	PR00599B 12.01 4.600e-20 173-189 PR00599A 11.65 1.273e-14 151-164 PR00599D 12.92 3.340e-10 273-285 PR00599C 11.34 6.471e-09 243-255
938	PD00289	PROTEIN SH3 DOMAIN REPEAT PRESYN.	PD00289 9.97 4.960e-10 137-150
940	PD02784	PROTEIN NUCLEAR RIBONUCLEOPROTEIN.	PD02784B 26.46 1.000e-40 217-259 PD02784C 20.76 1.000e-40 335-380 PD02784A 21.09 4.176e-36 178-214 PD02784B 26.46 7.683e-10 370-412
940	BL00030	Eukaryotic RNA-binding region RNP-1 proteins.	BL00030A 14.39 1.857e-09 456-474 BL00030A 14.39 1.000e-08 186-204
941	BL00740	MAM domain proteins.	BL00740A 13.87 7.188e-12 410-422
941	PR00020	MAM DOMAIN SIGNATURE	PR00020A 18.17 9.816e-12 408-426
941	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 4.082e-11 144-160
941	PF00094	von Willebrand factor type D domain proteins.	PF00094A 11.09 5.109e-09 139-148
941	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 7.632e-09 69-94
941	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 9.882e-09 146-172
941	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 9.100e-14 237-252 BL01187B 12.04 5.333e-12 192-207 BL01187B 12.04 6.333e-12 110-125 BL01187A 9.98 9.250e-09 173-184 BL01187A 9.98 1.000e-08 218-229
942	BL00740	MAM domain proteins.	BL00740A 13.87 7.188e-12 415-427
942	PR00020	MAM DOMAIN SIGNATURE	PR00020A 18.17 9.816e-12 413-431
942	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 4.082e-11 149-165
942	PF00094	von Willebrand factor type D domain proteins.	PF00094A 11.09 5.109e-09 144-153
942	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 7.632e-09 74-99
942	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 9.882e-09 151-177
942	BL01187	Calcium-binding EGF-like domain proteins	BL01187B 12.04 9.100e-14 242-257

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		pattern proteins.	BL01187B 12.04 5.333e-12 197-212 BL01187B 12.04 6.333e-12 115-130 BL01187A 9.98 9.250e-09 178-189 BL01187A 9.98 1.000e-08 223-234
943	PF00855	PWWP domain proteins.	PF00855 13.75 8.403e-13 274-290
943	BL00633	Bromodomain proteins.	BL00633B 13.82 8.977e-12 178-202
943	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 9.460e-10 94-109
943	PR00503	BROMODOMAIN SIGNATURE	PR00503B 9.96 8.667e-10 177-193 PR00503D 20.81 9.069e-09 211-230
944	PF00855	PWWP domain proteins.	PF00855 13.75 8.403e-13 274-290
944	BL00633	Bromodomain proteins.	BL00633B 13.82 8.977e-12 178-202
944	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 9.460e-10 94-109
944	PR00503	BROMODOMAIN SIGNATURE	PR00503B 9.96 8.667e-10 177-193 PR00503D 20.81 9.069e-09 211-230
945	PF00855	PWWP domain proteins.	PF00855 13.75 8.403e-13 274-290
945	BL00633	Bromodomain proteins.	BL00633B 13.82 8.977e-12 178-202
945	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 9.460e-10 94-109
945	PR00208	GLIADIN AND LMW GLUTENIN SUPERFAMILY SIGNATURE	PR00208A 12.59 9.868e-10 835-852 PR00208A 12.59 2.233e-09 838-855
945	DM00406	GLIADIN.	DM00406 7.73 9.000e-09 836-848
945	PR00503	BROMODOMAIN SIGNATURE	PR00503B 9.96 8.667e-10 177-193 PR00503D 20.81 9.069e-09 211-230
946	PF00855	PWWP domain proteins.	PF00855 13.75 8.403e-13 279-295
946	BL00633	Bromodomain proteins.	BL00633B 13.82 8.977e-12 183-207
946	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479B 12.57 9.460e-10 99-114
946	PR00208	GLIADIN AND LMW GLUTENIN SUPERFAMILY SIGNATURE	PR00208A 12.59 9.868e-10 840-857 PR00208A 12.59 2.233e-09 843-860
946	DM00406	GLIADIN.	DM00406 7.73 9.000e-09 841-853
946	PR00503	BROMODOMAIN SIGNATURE	PR00503B 9.96 8.667e-10 182-198 PR00503D 20.81 9.069e-09 216-235
950	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 4.039e-10 677-693
950	PR00206	CONNEXIN SIGNATURE	PR00206F 16.77 4.250e-09 498-521
950	PR00169	POTASSIUM CHANNEL SIGNATURE	PR00169G 9.39 7.932e-09 467-489
951	BL00427	Disintegrins proteins.	BL00427 13.93 7.592e-26 443-497
951	PR00138	MATRIXIN SIGNATURE	PR00138D 16.56 5.101e-11 342-367
951	BL00142	Neutral zinc metallopeptidases, zinc-binding region proteins.	BL00142 8.38 7.545e-11 342-352
951	PR00289	DISINTEGRIN SIGNATURE	PR00289A 13.62 2.500e-14 457-476 PR00289B 11.79 4.226e-10 486-498
951	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 8.909e-10 337-355
951	BL00546	Matrixins cysteine switch.	BL00546C 16.41 4.255e-09 336-367
951	BL00024	Hemopexin domain proteins.	BL00024D 17.28 5.596e-09 336-367
951	PR00907	THROMBOMODULIN SIGNATURE	PR00907E 11.70 7.353e-09 629-651
953	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 5.500e-11 360-372
953	PF00023	Ank repeat proteins.	PF00023A 16.03 6.000e-12 334-349 PF00023A 16.03 1.857e-11 156-171

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PF00023A 16.03 3.143e-11 255-270 PF00023B 14.20 3.455e-09 363-372 PF00023A 16.03 5.821e-09 188-203
953	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 4.273e-11 334-388 PF00791B 28.49 4.818e-11 301-355 PF00791B 28.49 4.845e-10 188-242 PF00791B 28.49 9.339e-09 222-276
954	BL00252	Interferon alpha, beta and delta family proteins.	BL00252A 18.49 6.657e-23 35-71 BL00252B 19.78 2.846e-14 73-123
954	PR00266	INTERFERON ALPHA AND BETA SUBUNIT SIGNATURE	PR00266A 13.61 1.000e-13 67-79
956	PR00081	GLUCOSE/RIBITOL DEHYDROGENASE FAMILY SIGNATURE	PR00081A 10.53 6.226e-13 34-51 PR00081F 15.71 7.632e-12 152-172 PR00081B 10.38 2.895e-10 108-119
958	PR00885	BACTERIAL GENERAL SECRETION PATHWAY PROTEIN H SIGNATURE	PR00885B 8.16 9.143e-10 394-408
958	BL00616	Histidine acid phosphatases phosphohistidine proteins.	BL00616A 11.86 7.811e-09 40-47
959	BL00284	Serpins proteins.	BL00284C 28.56 1.000e-34 118-159 BL00284D 16.34 4.857e-21 224-250 BL00284B 17.99 5.800e-19 91-111 BL00284E 19.15 7.577e-18 305-329
960	BL00284	Serpins proteins.	BL00284C 28.56 2.588e-23 186-227 BL00284A 15.64 7.750e-22 73-96 BL00284D 16.34 4.857e-21 292-318 BL00284E 19.15 7.577e-18 373-397
961	BL00284	Serpins proteins.	BL00284C 28.56 1.000e-34 186-227 BL00284A 15.64 7.750e-22 73-96 BL00284D 16.34 4.857e-21 292-318 BL00284B 17.99 6.625e-18 159-179 BL00284E 19.15 7.577e-18 373-397
962	BL00284	Serpins proteins.	BL00284C 28.56 1.000e-34 204-245 BL00284A 15.64 7.750e-22 73-96 BL00284B 17.99 5.800e-19 177-197 BL00284E 19.15 7.577e-18 373-397
964	BL00427	Disintegrins proteins.	BL00427 13.93 2.739e-16 459-513
964	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 9.045e-10 359-377
964	BL00142	Neutral zinc metallopeptidases, zinc-binding region proteins.	BL00142 8.38 1.429e-09 364-374
964	PR00289	DISINTEGRIN SIGNATURE	PR00289A 13.62 7.000e-14 473-492 PR00289B 11.79 2.579e-09 502-514
964	BL00412	Neuromodulin (GAP-43) proteins.	BL00412D 16.54 3.966e-11 763-813 BL00412D 16.54 7.065e-10 759-809 BL00412D 16.54 4.857e-09 764-814 BL00412D 16.54 9.357e-09 762-812
966	BL01238	GDA1/CD39 family of nucleoside phosphatases proteins.	BL01238C 14.36 2.174e-17 177-198 BL01238D 10.19 3.302e-13 216-229 BL01238A 11.72 6.936e-12 59-73 BL01238B 10.99 1.529e-09 133-143
967	BL01113	C1q domain proteins.	BL01113B 18.26 9.438e-20 95-130 BL01113D 7.47 9.308e-12 195-204

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL01113C 13.18 4.750e-10 163-182
967	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007B 14.16 7.698e-13 116-135 PR00007D 9.64 9.654e-11 193-203 PR00007C 15.60 3.656e-10 163-184 PR00007A 19.33 1.571e-09 89-115
969	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237A 11.48 5.355e-09 408-432
970	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 7.480e-10 160-182 BL00290B 13.17 2.875e-09 226-243
970	PR00939	C2HC-TYPE ZINC-FINGER SIGNATURE	PR00939B 13.27 8.412e-09 532-540
971	BL00289	Pentaxin family proteins.	BL00289D 17.60 1.947e-31 409-447 BL00289C 12.56 8.615e-16 370-388 BL00289A 30.36 7.457e-14 282-312 BL00289B 15.96 8.364e-12 327-341
971	PR00895	PENTAXIN SIGNATURE	PR00895E 12.74 5.065e-18 417-436 PR00895D 14.28 3.769e-17 397-416 PR00895C 12.29 4.273e-17 370-388 PR00895A 14.53 8.826e-13 305-319 PR00895B 14.20 2.154e-12 327-341 PR00895F 15.41 1.439e-10 436-450
972	PF00992	Troponin.	PF00992A 16.67 6.447e-09 741-775
973	BL00036	bZIP transcription factors basic domain proteins.	BL00036 9.02 5.737e-11 633-645
973	PR00043	JUN TRANSCRIPTION FACTOR SIGNATURE	PR00043B 8.73 9.241e-11 633-649
973	PF00624	Flocculin repeat proteins.	PF00624I 9.10 5.125e-10 461-490 PF00624I 9.10 5.800e-10 462-491 PF00624I 9.10 4.331e-09 458-487 PF00624I 9.10 6.457e-09 456-485 PF00624I 9.10 6.811e-09 453-482 PF00624I 9.10 8.441e-09 454-483
977	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 2.174e-10 2473-2486
977	DM00406	GLIADIN.	DM00406 7.73 1.400e-09 537-549
977	PR00021	SMALL PROLINE-RICH PROTEIN SIGNATURE	PR00021A 4.31 2.253e-09 538-550
977	BL00904	Protein prenyltransferases alpha subunit repeat proteins proteins.	BL00904A 8.30 2.660e-09 537-586
977	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 5.821e-10 543-575 DM00215 19.43 7.750e-10 531-563 DM00215 19.43 7.750e-10 559-591 DM00215 19.43 2.525e-09 536-568 DM00215 19.43 4.508e-09 533-565
977	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 9.017e-11 540-554 PR00049D 0.00 9.168e-11 541-555 PR00049D 0.00 2.983e-09 538-552 PR00049D 0.00 3.288e-09 539-553 PR00049D 0.00 3.898e-09 543-557 PR00049D 0.00 4.814e-09 537-551 PR00049D 0.00 6.034e-09 191-205
977	PR00239	MOLLUSCAN RHODOPSIN C-TERMINAL TAIL SIGNATURE	PR00239E 1.58 6.318e-09 542-553

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
977	BL00415	Synapsins proteins.	BL00415N 4.29 8.143e-11 556-599 BL00415N 4.29 8.357e-11 550-593 BL00415N 4.29 6.702e-10 543-586 BL00415N 4.29 8.145e-10 532-575 BL00415N 4.29 8.969e-10 548-591 BL00415N 4.29 3.562e-09 555-598 BL00415N 4.29 4.088e-09 531-574 BL00415N 4.29 9.869e-09 539-582
977	PR00211	GLUTELIN SIGNATURE	PR00211B 0.86 9.917e-09 551-571
980	BL00282	Kazal serine protease inhibitors family proteins.	BL00282 16.88 4.234e-12 73-95
980	PR00834	HTRA/DEGQ PROTEASE FAMILY SIGNATURE	PR00834C 15.43 3.613e-20 237-261 PR00834D 12.14 6.455e-18 275-292 PR00834B 10.09 5.500e-14 196-216 PR00834E 13.63 5.355e-13 297-314 PR00834F 10.91 9.526e-12 389-401 PR00834A 9.80 3.659e-11 175-187
980	BL00222	Insulin-like growth factor binding proteins.	BL00222B 11.09 4.420e-10 22-37
980	PR00290	KAZAL-TYPE SERINE PROTEASE INHIBITOR SIGNATURE	PR00290B 9.78 4.326e-09 84-95
980	BL00273	Heat-stable enterotoxins proteins.	BL00273 12.24 8.286e-09 26-38
981	PR00792	PEPSIN (A1) ASPARTIC PROTEASE FAMILY SIGNATURE	PR00792A 11.54 5.500e-18 80-100 PR00792D 12.74 9.069e-13 395-410 PR00792C 9.10 4.214e-12 312-323
981	BL00141	Eukaryotic and viral aspartyl proteases proteins.	BL00141A 12.10 4.789e-15 87-102 BL00141E 14.32 6.850e-15 396-419 BL00141D 6.28 7.300e-11 312-321 BL00141B 12.14 2.929e-10 228-239
982	BL00523	Sulfatases proteins.	BL00523A 13.36 6.651e-10 44-60
984	PR00765	CARBOXYPEPTIDASE A METALLOPROTEASE (M14) FAMILY SIGNATURE	PR00765B 15.57 7.857e-16 99-113 PR00765D 14.16 5.500e-11 233-246 PR00765C 12.55 1.290e-10 179-187
984	BL00132	Zinc carboxypeptidases, zinc-binding region 1 proteins.	BL00132C 21.35 3.308e-28 129-169 BL00132B 15.93 1.871e-16 99-112 BL00132A 26.07 1.682e-14 50-90 BL00132F 13.26 7.254e-14 228-249 BL00132D 12.70 2.875e-12 173-187 BL00132E 17.72 3.552e-12 199-225 BL00132G 10.94 4.541e-10 285-302
985	PR00765	CARBOXYPEPTIDASE A METALLOPROTEASE (M14) FAMILY SIGNATURE	PR00765B 15.57 7.857e-16 99-113 PR00765D 14.16 5.500e-11 233-246 PR00765C 12.55 1.290e-10 179-187
985	BL00132	Zinc carboxypeptidases, zinc-binding region 1 proteins.	BL00132C 21.35 3.308e-28 129-169 BL00132B 15.93 1.871e-16 99-112 BL00132A 26.07 1.682e-14 50-90 BL00132F 13.26 7.254e-14 228-249 BL00132D 12.70 2.875e-12 173-187 BL00132E 17.72 3.552e-12 199-225 BL00132G 10.94 4.541e-10 285-302
990	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 5.304e-11 110-122
991	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 1.000e-15 139-169

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00107B 13.31 4.273e-13 209-224
991	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 7.894e-13 139-157
991	BL00240	Receptor tyrosine kinase class III proteins.	BL00240E 11.56 6.580e-10 125-162
994	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007A 19.33 6.936e-13 168-194 PR00007C 15.60 9.250e-13 243-264 PR00007B 14.16 9.372e-13 195-214 PR00007D 9.64 5.500e-11 275-285
994	PR00524	CHOLECYSTOKININ TYPE A RECEPTOR SIGNATURE	PR00524F 5.36 1.766e-09 94-107
994	BL00420	Speract receptor repeat proteins domain proteins.	BL00420A 20.42 7.058e-12 79-107 BL00420A 20.42 4.689e-10 97-125 BL00420A 20.42 6.902e-10 82-110 BL00420A 20.42 1.277e-09 85-113 BL00420A 20.42 5.292e-09 76-104
994	BL01113	C1q domain proteins.	BL01113B 18.26 1.675e-24 174-209 BL01113A 17.99 1.871e-15 85-111 BL01113A 17.99 5.091e-14 82-108 BL01113D 7.47 3.250e-13 277-286 BL01113A 17.99 4.892e-13 76-102 BL01113A 17.99 6.108e-13 94-120 BL01113A 17.99 9.757e-13 79-105 BL01113A 17.99 3.769e-12 88-114 BL01113A 17.99 6.308e-12 91-117 BL01113C 13.18 9.294e-12 243-262 BL01113A 17.99 8.159e-11 70-96 BL01113A 17.99 9.795e-11 97-123 BL01113A 17.99 9.809e-10 73-99 BL01113A 17.99 6.019e-09 103-129
995	DM01595	kw ALLANTOICASE SPAC1F7.09C.	DM01595D 10.94 8.269e-16 116-140 DM01595I 8.91 2.714e-15 300-317 DM01595I 8.91 9.727e-14 117-134 DM01595D 10.94 3.274e-11 299-323 DM01595E 14.67 6.299e-09 152-184
997	BL00720	Guanine-nucleotide dissociation stimulators CDC25 family sign.	BL00720B 16.57 4.103e-18 1089-1112
997	BL00741	Guanine-nucleotide dissociation stimulators CDC24 family sign.	BL00741B 14.27 4.326e-16 377-399
1001	BL00048	Protamine P1 proteins.	BL00048 6.39 6.684e-10 949-975 BL00048 6.39 3.363e-09 947-973 BL00048 6.39 9.888e-09 781-807
1002	PF00628	PHD-finger.	PF00628 15.84 8.412e-14 201-215
1002	BL00048	Protamine P1 proteins.	BL00048 6.39 6.684e-10 1158-1184 BL00048 6.39 3.363e-09 1156-1182 BL00048 6.39 9.888e-09 990-1016
1003	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 4.103e-11 1132-1146 PR00320C 13.01 8.200e-10 1132-1146 PR00320A 16.74 9.735e-10 1091-1105 PR00320C 13.01 2.500e-09 1091-1105 PR00320B 12.19 6.625e-09 1132-1146
1004	PF00569	Zinc finger present in dystrophin, CBP/p300.	PF00569 13.42 1.545e-16 21-37

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1004	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306A 10.26 2.929e-09 257-270
1006	PR00399	SYNAPTOTAGMIN SIGNATURE	PR00399A 9.52 1.964e-09 162-177
1007	PR00806	VINCULIN SIGNATURE	PR00806D 11.95 3.963e-09 564-579
1008	BL00319	Amyloidogenic glycoprotein extracellular domain proteins.	BL00319C 17.12 5.625e-10 565-598 BL00319C 17.12 4.316e-09 563-596 BL00319C 17.12 5.382e-09 560-593
1008	PF00922	Vesiculovirus phosphoprotein.	PF00922A 19.17 8.862e-09 571-604
1009	PR00405	HIV REV INTERACTING PROTEIN SIGNATURE	PR00405B 11.83 8.385e-15 281-298 PR00405A 17.71 4.306e-14 262-281
1009	PR00452	SH3 DOMAIN SIGNATURE	PR00452B 11.65 5.500e-09 895-910
1009	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 9.036e-09 335-347
1011	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 2.674e-10 384-407 BL00240B 24.70 8.535e-10 479-502 BL00240B 24.70 7.702e-09 575-598
1011	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 4.600e-10 617-649 PD02870B 18.83 5.883e-09 28-60 PD02870B 18.83 7.894e-09 225-257
1015	BL00018	EF-hand calcium-binding domain proteins.	BL00018 7.41 5.765e-11 147-159
1015	PR00450	RECOVERIN FAMILY SIGNATURE	PR00450C 12.22 1.228e-09 33-54
1015	BL00303	S-100/ICaBP type calcium binding protein.	BL00303B 26.15 6.559e-09 26-62
1018	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 1.474e-24 136-175 BL00237C 13.19 6.400e-14 289-315 BL00237B 5.28 3.077e-12 244-255 BL00237D 11.23 9.654e-11 342-358
1018	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237E 13.03 2.588e-16 236-259 PR00237D 8.94 8.800e-14 186-207 PR00237B 13.50 2.636e-13 105-126 PR00237C 15.69 4.960e-13 150-172 PR00237F 13.57 6.040e-13 294-318 PR00237A 11.48 3.143e-12 72-96 PR00237G 19.63 3.531e-12 332-358 PR00237E 13.03 4.441e-09 234-257
1018	PR00238	OPSIN SIGNATURE	PR00238B 16.24 2.667e-14 208-220 PR00238A 13.79 8.286e-09 93-105
1018	PR00667	RETINAL PIGMENT EPITHELIUM-RETINAL GPCR SIGNATURE	PR00667B 10.86 8.800e-09 91-106
1019	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 5.500e-15 378-391 PR00019A 11.19 3.739e-10 134-147 PR00019B 11.36 1.000e-09 535-548 PR00019B 11.36 2.440e-09 375-388 PR00019A 11.19 3.333e-09 252-265 PR00019B 11.36 4.960e-09 225-238 PR00019A 11.19 7.000e-09 560-573 PR00019B 11.36 7.840e-09 351-364 PR00019B 11.36 9.640e-09 180-193
1021	BL00720	Guanine-nucleotide dissociation stimulators CDC25 family sign.	BL00720B 16.57 6.595e-15 996-1019
1021	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791C 20.98 6.011e-12 606-644

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1021	PD00289	PROTEIN SH3 DOMAIN REPEAT PRESYN.	PD00289 9.97 5.050e-11 625-638
1021	PR00834	HTRA/DEGQ PROTEASE FAMILY SIGNATURE	PR00834F 10.91 2.946e-09 621-633
1021	BL00888	Cyclic nucleotide-binding domain proteins.	BL00888B 14.79 4.682e-09 355-378
1022	BL00720	Guanine-nucleotide dissociation stimulators CDC25 family sign.	BL00720B 16.57 6.595e-15 946-969
1022	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791C 20.98 6.011e-12 556-594
1022	PD00289	PROTEIN SH3 DOMAIN REPEAT PRESYN.	PD00289 9.97 5.050e-11 575-588
1022	PR00834	HTRA/DEGQ PROTEASE FAMILY SIGNATURE	PR00834F 10.91 2.946e-09 571-583
1022	BL00888	Cyclic nucleotide-binding domain proteins.	BL00888B 14.79 4.682e-09 305-328
1024	BL00476	Fatty acid desaturases family 1 proteins.	BL00476B 18.34 5.420e-09 365-408
1024	PR00669	INHIBIN ALPHA CHAIN SIGNATURE	PR00669B 8.27 6.488e-09 204-220
1025	BL00476	Fatty acid desaturases family 1 proteins.	BL00476B 18.34 5.420e-09 327-370
1025	PR00669	INHIBIN ALPHA CHAIN SIGNATURE	PR00669B 8.27 6.488e-09 166-182
1028	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 9.419e-36 133-180 BL00232B 32.79 5.345e-21 242-289 BL00232A 27.72 3.727e-20 39-71 BL00232C 10.65 2.742e-14 240-257 BL00232B 32.79 6.566e-14 357-404
1028	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 2.909e-15 240-257 PR00205A 14.73 8.457e-11 165-180
1029	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 9.419e-36 133-180 BL00232B 32.79 5.345e-21 242-289 BL00232A 27.72 3.727e-20 39-71 BL00232C 10.65 2.742e-14 240-257 BL00232B 32.79 6.566e-14 357-404
1029	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 2.909e-15 240-257 PR00205A 14.73 8.457e-11 165-180
1030	PF00816	H-NS histone family.	PF00816B 13.84 9.284e-09 102-131
1030	PR00124	ATP SYNTHASE C SUBUNIT SIGNATURE	PR00124A 8.81 9.000e-10 41-60 PR00124A 8.81 9.379e-09 43-62
1030	BL00604	Synaptophysin / synaptoporin proteins.	BL00604F 5.96 9.696e-09 41-85
1031	BL00869	Renal dipeptidase proteins.	BL00869C 12.58 3.172e-19 112-147 BL00869E 13.12 9.129e-18 173-209 BL00869J 15.60 6.032e-17 323-362 BL00869H 11.08 1.840e-16 272-294 BL00869G 13.55 2.543e-16 245-266 BL00869F 12.77 7.031e-14 210-244 BL00869I 12.92 3.274e-12 295-322 BL00869D 14.02 5.282e-10 148-176 BL00869B 15.55 9.382e-10 84-113
1032	BL00218	Amino acid permeases proteins.	BL00218D 21.49 7.446e-11 244-288 BL00218E 23.30 3.640e-10 325-364
1033	BL00721	Formate--tetrahydrofolate ligase proteins.	BL00721B 13.21 1.000e-40 456-510 BL00721D 13.90 1.000e-40 648-701 BL00721E 13.46 1.000e-40 707-755 BL00721I 18.79 2.500e-40 924-969

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00721H 21.20 8.239e-39 873-923 BL00721A 15.31 9.719e-32 397-430 BL00721C 16.92 4.000e-30 608-644 BL00721F 15.96 8.232e-27 770-811 BL00721G 7.97 3.017e-10 831-843
1033	PR00085	TETRAHYDROFOLATE DEHYDROGENASE/CYCLOHYDROLASE FAMILY SIGNATURE	PR00085C 15.23 4.906e-15 169-190 PR00085B 15.92 7.488e-10 136-163 PR00085E 15.79 6.216e-09 266-295
1033	BL00415	Synapsins proteins.	BL00415N 4.29 8.489e-09 18-61
1035	PR00834	HTRA/DEGQ PROTEASE FAMILY SIGNATURE	PR00834F 10.91 2.946e-09 82-94
1035	BL00741	Guanine-nucleotide dissociation stimulators CDC24 family sign.	BL00741B 14.27 2.962e-09 911-933
1035	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 4.814e-09 1125-1139 PR00049D 0.00 5.729e-09 147-161
1035	PR00554	ADENOSINE A2B RECEPTOR SIGNATURE	PR00554B 12.52 8.855e-09 724-732
1037	PR00390	PHOSPHOLIPASE C SIGNATURE	PR00390A 15.09 1.439e-20 295-313
1037	BL00303	S-100/ICaBP type calcium binding protein.	BL00303B 26.15 4.971e-09 135-171
1037	BL00292	Cyclins proteins.	BL00292A 22.87 5.114e-09 220-253
1039	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245B 10.38 5.821e-14 176-190 PR00245A 18.03 6.891e-14 58-79 PR00245E 12.40 6.170e-11 290-304 PR00245C 7.84 2.286e-10 237-252
1039	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 5.408e-09 89-128
1039	PR00896	VASOPRESSIN RECEPTOR SIGNATURE	PR00896B 9.01 7.577e-09 54-65
1039	PR00534	MELANOCORTIN RECEPTOR FAMILY SIGNATURE	PR00534A 11.49 8.586e-09 50-62
1039	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237B 13.50 6.000e-09 58-79 PR00237E 13.03 8.941e-09 198-221
1040	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187A 9.98 2.125e-12 233-244 BL01187A 9.98 4.789e-11 286-297 BL01187B 12.04 3.057e-10 348-363
1040	PD00919	CALCIUM-BINDING PRECURSOR SIGNAL R.	PD00919D 17.80 1.000e-40 406-456 PD00919D 17.80 1.000e-40 465-515 PD00919G 15.92 1.000e-40 590-633 PD00919H 17.48 1.000e-40 634-675 PD00919I 18.44 1.000e-40 676-724 PD00919J 16.09 1.000e-40 737-775 PD00919K 18.26 1.000e-40 776-810 PD00919L 16.90 1.000e-40 812-851 PD00919C 12.28 9.250e-34 357-386 PD00919F 11.63 7.000e-33 555-583 PD00919E 11.16 1.000e-32 521-549 PD00919G 15.92 4.197e-23 453-496 PD00919G 15.92 1.556e-20 394-437 PD00919F 11.63 5.103e-20 399-427 PD00919G 15.92 9.111e-20 550-593 PD00919D 17.80 3.793e-19 526-576 PD00919F 11.63 8.397e-18 458-486 PD00919B 9.47 3.455e-17 308-322

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PD00919D 17.80 6.967e-17 566-616 PD00919A 11.53 3.520e-15 199-210 PD00919F 11.63 6.000e-15 595-623 PD00919D 17.80 3.970e-14 488-538 PD00919D 17.80 8.110e-14 429-479 PD00919F 11.63 3.379e-13 517-545 PD00919G 15.92 4.757e-12 489-532 PD00919D 17.80 6.094e-12 370-420 PD00919D 17.80 9.915e-12 562-612 PD00919E 11.16 2.517e-11 403-431 PD00919B 9.47 3.714e-11 215-229 PD00919G 15.92 7.224e-11 512-555 PD00919F 11.63 8.372e-11 494-522 PD00919E 11.16 8.382e-11 498-526 PD00919E 11.16 9.899e-11 462-490 PD00919E 11.16 2.663e-10 559-587 PD00919D 17.80 9.061e-10 501-551 PD00919E 11.16 1.092e-09 599-627 PD00919D 17.80 1.525e-09 503-553 PD00919G 15.92 3.638e-09 430-473 PD00919E 11.16 4.582e-09 439-467 PD00919D 17.80 6.625e-09 524-574 PD00919A 11.53 6.727e-09 239-250 PD00919D 17.80 6.775e-09 442-492
1042	BL01022	PTR2 family proton/oligopeptide symporters proteins.	BL01022B 22.19 2.241e-15 74-119 BL01022E 23.51 3.739e-14 440-475 BL01022A 11.58 2.212e-12 44-62 BL01022D 9.42 2.946e-12 195-207 BL01022C 16.62 6.226e-10 160-183
1042	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308C 3.83 2.169e-09 20-29
1043	PF01140	Matrix protein (MA), p15.	PF01140D 15.54 3.700e-10 977-1011
1043	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 5.018e-10 542-574 DM00215 19.43 8.322e-09 537-569 DM00215 19.43 8.322e-09 541-573 DM00215 19.43 8.627e-09 530-562 DM00215 19.43 9.542e-09 540-572
1044	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 9.727e-36 10-48
1044	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 3.769e-15 384-396 PD00066 13.92 4.462e-15 244-256 PD00066 13.92 6.538e-15 468-480 PD00066 13.92 1.000e-13 300-312 PD00066 13.92 1.000e-13 608-620 PD00066 13.92 9.000e-13 160-172 PD00066 13.92 3.571e-12 216-228 PD00066 13.92 4.000e-12 580-592 PD00066 13.92 5.714e-12 496-508 PD00066 13.92 2.957e-11 524-536 PD00066 13.92 7.652e-11 328-340 PD00066 13.92 2.385e-10 552-564 PD00066 13.92 1.600e-09 272-284

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1044	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 2.636e-15 589-602 PR00048A 10.52 4.273e-15 253-266 PR00048A 10.52 5.500e-14 533-546 PR00048A 10.52 4.214e-13 225-238 PR00048A 10.52 5.765e-12 281-294 PR00048A 10.52 7.882e-12 477-490 PR00048A 10.52 1.474e-11 169-182 PR00048A 10.52 1.947e-11 141-154 PR00048A 10.52 3.368e-11 309-322 PR00048A 10.52 8.105e-11 561-574 PR00048A 10.52 9.526e-11 393-406 PR00048B 6.02 1.000e-10 297-306 PR00048B 6.02 1.563e-10 577-586 PR00048B 6.02 3.250e-10 353-362 PR00048B 6.02 3.250e-10 409-418 PR00048B 6.02 3.250e-10 437-446 PR00048A 10.52 4.522e-10 617-630 PR00048B 6.02 4.938e-10 241-250 PR00048B 6.02 7.750e-10 493-502 PR00048B 6.02 8.875e-10 381-390 PR00048B 6.02 8.875e-10 465-474 PR00048A 10.52 2.440e-09 197-210 PR00048B 6.02 4.789e-09 605-614
1044	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 7.429e-16 536-552 BL00028 16.07 2.125e-15 592-608 BL00028 16.07 4.938e-15 256-272 BL00028 16.07 5.950e-13 228-244 BL00028 16.07 1.000e-11 452-468 BL00028 16.07 2.731e-11 396-412 BL00028 16.07 4.115e-11 172-188 BL00028 16.07 5.154e-11 284-300 BL00028 16.07 5.846e-11 480-496 BL00028 16.07 6.538e-11 564-580 BL00028 16.07 9.654e-11 620-636 BL00028 16.07 1.300e-10 144-160 BL00028 16.07 1.900e-10 340-356 BL00028 16.07 1.900e-10 424-440 BL00028 16.07 9.100e-10 116-132 BL00028 16.07 9.100e-10 200-216 BL00028 16.07 9.700e-10 368-384 BL00028 16.07 5.629e-09 508-524 BL00028 16.07 7.943e-09 312-328
1046	PD01795	PROTEIN AMINOPEPTIDASE PRECURSOR HYDROLASE SIGNA.	PD01795A 10.27 6.667e-09 362-370
1049	PF00651	BTB (also known as BR-C/Ttk) domain proteins.	PF00651 15.00 7.840e-09 43-55
1049	PR00766	AMILORIDE-SENSITIVE AMINE OXIDASE SIGNATURE	PR00766G 11.62 9.905e-09 91-111
1050	BL00211	ABC transporters family proteins.	BL00211B 13.37 7.429e-20 141-172
1051	PF00569	Zinc finger present in dystrophin, CBP/p300.	PF00569 13.42 1.545e-16 21-37
1051	PD00306	PROTEIN GLYCOPROTEIN PRECURSOR RE.	PD00306A 10.26 2.929e-09 257-270

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1052	DM00031	IMMUNOGLOBULIN V REGION.	DM00031B 15.41 5.500e-12 77-110
1052	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 9.100e-12 154-176
1053	PR00018	KRINGLE DOMAIN SIGNATURE	PR00018A 14.52 3.423e-09 36-51
1054	DM01688	2 POLY-IG RECEPTOR.	DM01688B 15.06 4.504e-09 85-132 DM01688J 14.69 8.364e-09 32-68
1055	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290B 13.17 4.000e-21 281-298 BL00290A 20.89 4.600e-16 34-56 BL00290A 20.89 4.375e-15 224-246
1064	PD00289	PROTEIN SH3 DOMAIN REPEAT PRESYN.	PD00289 9.97 1.000e-09 453-466 PD00289 9.97 5.034e-09 47-60 PD00289 9.97 5.034e-09 258-271
1064	PF00595	PDZ domain proteins (Also known as DHR or GLGF).	PF00595 13.40 9.250e-10 450-460 PF00595 13.40 7.000e-09 255-265
1067	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 9.471e-12 126-144
1067	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 2.800e-22 126-156 BL00107B 13.31 6.786e-11 196-211
1067	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479C 12.01 3.000e-09 174-186
1067	BL00790	Receptor tyrosine kinase class V proteins.	BL00790M 8.74 4.857e-09 117-138
1068	PR00179	LIPOCALIN SIGNATURE	PR00179B 9.56 1.000e-12 120-132 PR00179C 19.02 1.000e-10 148-163 PR00179A 13.78 5.680e-10 37-49
1068	BL00213	Lipocalin proteins.	BL00213B 8.78 8.000e-10 120-130 BL00213A 12.95 9.526e-10 37-50
1070	PR00200	ANNEXIN TYPE IV SIGNATURE	PR00200G 9.43 5.602e-17 299-325 PR00200E 10.00 6.160e-16 136-157 PR00200E 10.00 3.012e-13 295-316 PR00200F 13.72 6.157e-13 219-245 PR00200E 10.00 4.742e-12 64-85 PR00200B 7.39 9.063e-12 69-91 PR00200G 9.43 1.991e-11 140-166 PR00200D 10.01 5.304e-11 109-125 PR00200H 13.68 5.050e-10 343-356 PR00200B 7.39 2.865e-09 141-163
1070	PR00202	ANNEXIN TYPE VI SIGNATURE	PR00202G 8.01 1.563e-14 299-325 PR00202E 13.00 9.613e-13 219-245 PR00202D 5.58 8.636e-11 136-157 PR00202G 8.01 2.525e-09 140-166 PR00202D 5.58 3.560e-09 64-85
1070	PR00199	ANNEXIN TYPE III SIGNATURE	PR00199F 16.19 7.387e-18 219-245 PR00199D 5.65 1.409e-16 295-316 PR00199G 9.09 6.354e-16 300-325 PR00199D 5.65 6.455e-16 136-157 PR00199D 5.65 1.474e-13 64-85 PR00199B 6.86 2.346e-10 69-91 PR00199B 6.86 5.458e-10 300-322 PR00199B 6.86 8.234e-10 141-163 PR00199C 13.84 6.464e-09 109-125
1070	PR00197	ANNEXIN TYPE I SIGNATURE	PR00197D 7.50 5.629e-16 136-157

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00197F 9.03 7.395e-15 299-319 PR00197D 7.50 1.234e-14 295-316 PR00197E 11.89 3.541e-13 219-245 PR00197D 7.50 6.379e-11 64-85 PR00197B 7.56 7.124e-09 69-91
1070	PR00198	ANNEXIN TYPE II SIGNATURE	PR00198D 7.65 2.222e-15 136-157 PR00198D 7.65 3.647e-13 295-316 PR00198G 8.09 4.375e-13 299-319 PR00198D 7.65 9.165e-10 64-85 PR00198B 8.71 7.529e-09 69-91 PR00198C 14.32 7.900e-09 109-125 PR00198G 8.09 8.125e-09 140-160
1070	BL00223	Annexins repeat proteins domain proteins.	BL00223C 24.79 1.000e-40 278-332 BL00223B 28.47 9.679e-39 201-250 BL00223A 15.59 1.000e-27 132-165 BL00223A 15.59 6.936e-22 60-93 BL00223C 24.79 3.077e-17 119-173 BL00223A 15.59 4.194e-16 291-324 BL00223C 24.79 2.514e-09 47-101 BL00223B 28.47 8.533e-09 117-166
1070	PR00201	ANNEXIN TYPE V SIGNATURE	PR00201G 11.02 7.692e-19 299-325 PR00201D 10.49 1.656e-11 136-157 PR00201A 6.05 6.242e-11 69-91 PR00201E 12.37 8.040e-11 219-245 PR00201C 11.13 3.897e-10 109-125 PR00201D 10.49 5.050e-10 64-85 PR00201G 11.02 6.215e-10 140-166 PR00201D 10.49 9.910e-10 295-316 PR00201A 6.05 4.297e-09 300-322 PR00201H 12.04 7.506e-09 343-356 PR00201A 6.05 8.842e-09 141-163
1070	PR00196	ANNEXIN FAMILY SIGNATURE	PR00196D 21.86 2.895e-21 219-245 PR00196E 9.19 3.077e-20 299-319 PR00196C 10.36 5.500e-20 136-157 PR00196A 11.16 7.632e-19 69-91 PR00196C 10.36 1.500e-15 295-316 PR00196B 10.68 8.875e-15 109-125 PR00196C 10.36 8.071e-14 64-85 PR00196A 11.16 2.714e-12 141-163 PR00196G 11.72 4.250e-12 343-356 PR00196E 9.19 9.735e-12 140-160 PR00196F 13.89 1.000e-11 327-342 PR00196A 11.16 8.859e-10 300-322 PR00196F 13.89 7.938e-09 168-183 PR00196D 21.86 9.775e-09 135-161
1071	BL00610	Sodium:neurotransmitter symporter family proteins.	BL00610A 17.73 1.000e-40 52-101 BL00610B 23.65 1.000e-40 115-164 BL00610C 12.94 1.000e-40 212-263 BL00610E 20.34 1.000e-40 372-414 BL00610F 29.02 1.000e-40 469-523 BL00610G 12.89 9.217e-22 528-550 BL00610D 20.97 4.822e-19 278-330

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1071	PR00176	SODIUM/NEUROTRANSMITTER SYMPORTER SIGNATURE	PR00176A 16.82 1.529e-26 52-73 PR00176C 10.84 5.500e-25 124-150 PR00176G 12.48 2.688e-22 458-478 PR00176E 11.41 2.000e-21 322-342 PR00176F 10.73 3.333e-20 376-395 PR00176B 7.31 1.600e-19 81-100 PR00176D 9.02 1.321e-18 239-256 PR00176H 15.27 2.440e-18 498-518
1072	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 7.652e-09 113-122
1073	BL01207	Glypicans proteins.	BL01207C 19.08 6.538e-31 250-285 BL01207B 23.69 9.122e-28 191-236 BL01207D 23.23 1.692e-24 429-463 BL01207A 12.21 1.000e-16 62-77 BL01207E 13.70 1.214e-11 487-503
1073	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 3.898e-09 515-529
1073	BL00291	Prion protein.	BL00291A 4.49 7.724e-09 530-564
1073	PR00829	MAJOR POLLEN ALLERGEN LOL PI FAMILY SIGNATURE	PR00829E 10.81 9.597e-09 306-320
1075	PF00777	Sialyltransferase family.	PF00777C 18.60 2.581e-28 294-348
1078	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 4.541e-13 790-816
1078	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477J 19.04 3.382e-27 1241-1271 BL00477F 17.34 8.500e-25 785-814 BL00477G 19.43 8.826e-23 983-1014 BL00477A 13.50 9.800e-23 122-150 BL00477L 23.51 5.500e-16 1437-1469 BL00477K 17.42 4.529e-14 1382-1405 BL00477E 17.53 6.538e-13 755-775 BL00477B 9.05 6.625e-13 209-221 BL00477I 18.76 2.650e-12 1085-1111 BL00477D 12.73 4.073e-12 729-738 BL00477H 9.07 5.395e-12 1054-1065 BL00477C 15.70 1.161e-10 236-252
1079	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 4.541e-13 804-830
1079	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477F 17.34 8.500e-25 799-828 BL00477A 13.50 9.800e-23 135-163 BL00477E 17.53 6.538e-13 769-789 BL00477B 9.05 6.625e-13 222-234 BL00477D 12.73 4.073e-12 743-752 BL00477C 15.70 1.161e-10 249-265
1080	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477A 13.50 9.800e-23 122-150 BL00477B 9.05 6.625e-13 209-221 BL00477C 15.70 1.161e-10 236-252
1081	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 4.541e-13 790-816
1081	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477J 19.04 3.382e-27 1241-1271 BL00477F 17.34 8.500e-25 785-814 BL00477G 19.43 8.826e-23 983-1014 BL00477A 13.50 9.800e-23 122-150 BL00477L 23.51 8.800e-22 1437-1469 BL00477K 17.42 4.529e-14 1382-1405 BL00477E 17.53 6.538e-13 755-775 BL00477B 9.05 6.625e-13 209-221

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00477I 18.76 2.650e-12 1085-1111 BL00477D 12.73 4.073e-12 729-738 BL00477H 9.07 5.395e-12 1054-1065 BL00477C 15.70 1.161e-10 236-252
1081	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115V 21.32 5.745e-09 1422-1471
1082	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 4.541e-13 791-817
1082	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477F 17.34 8.500e-25 786-815 BL00477A 13.50 9.800e-23 122-150 BL00477E 17.53 6.538e-13 756-776 BL00477B 9.05 6.625e-13 209-221 BL00477D 12.73 4.073e-12 730-739 BL00477C 15.70 1.161e-10 236-252
1083	BL00122	Carboxylesterases type-B serine proteins.	BL00122E 22.02 9.027e-31 195-235 BL00122A 12.04 5.500e-16 60-80 BL00122D 12.53 7.545e-16 171-186 BL00122C 7.91 8.125e-13 142-152 BL00122B 16.84 4.830e-10 122-132 BL00122F 11.10 5.500e-10 247-256 BL00122G 11.67 9.625e-10 500-510
1083	PR00878	CHOLINESTERASE SIGNATURE	PR00878F 5.37 7.171e-09 460-472
1084	PD00919	CALCIUM-BINDING PRECURSOR SIGNAL R.	PD00919B 9.47 7.485e-10 1019-1033
1084	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 9.138e-10 175-220
1084	BL00279	Membrane attack complex components / perforin proteins.	BL00279E 37.11 9.241e-10 387-434
1084	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011D 14.03 2.696e-09 413-431
1084	PR00907	THROMBOMODULIN SIGNATURE	PR00907G 11.63 7.973e-09 890-916
1084	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 8.017e-09 92-106
1084	BL00022	EGF-like domain proteins.	BL00022B 7.54 8.200e-09 1187-1193
1084	PR00010	TYPE II EGF-LIKE SIGNATURE	PR00010C 11.16 7.667e-11 1183-1193 PR00010C 11.16 1.857e-09 937-947 PR00010C 11.16 4.857e-09 1687-1697 PR00010C 11.16 8.286e-09 1642-1652
1084	PR00009	TYPE I EGF SIGNATURE	PR00009C 14.11 9.118e-09 1058-1069
1084	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 7.000e-17 1682-1697 BL01187B 12.04 2.350e-14 1178-1193 BL01187B 12.04 5.500e-14 1136-1151 BL01187B 12.04 1.391e-13 642-657 BL01187B 12.04 4.130e-13 1219-1234 BL01187B 12.04 4.913e-13 1095-1110 BL01187B 12.04 9.609e-13 932-947 BL01187B 12.04 9.667e-12 1054-1069 BL01187B 12.04 4.600e-11 1261-1276 BL01187A 9.98 9.526e-11 997-1008 BL01187B 12.04 1.257e-10 1483-1498 BL01187A 9.98 7.857e-10 1078-1089 BL01187A 9.98 2.875e-09 1243-1254 BL01187B 12.04 3.250e-09 1637-1652 BL01187A 9.98 7.000e-09 914-925 BL01187A 9.98 1.000e-08 1037-1048

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1086	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014A 8.22 8.941e-10 816-825 PR00014D 12.04 5.950e-09 872-886 PR00014C 15.44 6.478e-09 854-872
1086	BL00790	Receptor tyrosine kinase class V proteins.	BL00790I 20.01 6.250e-12 865-895 BL00790I 20.01 7.750e-09 662-692
1087	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 2.737e-24 16-54
1087	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 4.150e-13 219-235 BL00028 16.07 7.300e-13 191-207 BL00028 16.07 4.522e-12 163-179 BL00028 16.07 2.038e-11 247-263
1087	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 7.231e-15 235-247 PD00066 13.92 6.143e-12 179-191 PD00066 13.92 7.923e-10 207-219
1087	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 3.250e-14 188-201 PR00048A 10.52 4.000e-14 244-257 PR00048A 10.52 4.706e-12 216-229 PR00048B 6.02 3.250e-10 232-241 PR00048A 10.52 2.440e-09 160-173 PR00048B 6.02 9.053e-09 260-269
1088	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 2.737e-24 16-54
1088	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 8.043e-12 163-179
1088	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 2.800e-09 160-173 PR00048B 6.02 9.053e-09 176-185
1089	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 1.127e-09 86-128 BL00243I 31.77 2.775e-09 30-72 BL00243I 31.77 5.437e-09 89-131
1089	BL01208	VWFC domain proteins.	BL01208B 15.83 5.865e-09 114-128
1089	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 5.613e-09 24-51 PD02283C 17.54 5.613e-09 68-95 PD02283C 17.54 7.188e-09 93-120 PD02283C 17.54 7.750e-09 103-130
1089	BL00269	Mammalian defensins proteins.	BL00269C 16.52 9.289e-09 28-56 BL00269C 16.52 9.289e-09 72-100
1089	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 6.897e-12 66-111 BL00203 13.94 3.769e-11 70-115 BL00203 13.94 4.165e-11 40-85 BL00203 13.94 6.835e-11 65-110 BL00203 13.94 1.096e-10 61-106 BL00203 13.94 2.723e-10 21-66 BL00203 13.94 2.723e-10 22-67 BL00203 13.94 5.213e-10 91-136 BL00203 13.94 5.883e-10 26-71 BL00203 13.94 7.032e-10 114-159 BL00203 13.94 1.643e-09 85-130 BL00203 13.94 1.735e-09 105-150 BL00203 13.94 2.745e-09 80-125 BL00203 13.94 3.388e-09 56-101 BL00203 13.94 4.214e-09 81-126 BL00203 13.94 5.500e-09 60-105

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00203 13.94 6.694e-09 100-145 BL00203 13.94 6.969e-09 17-62 BL00203 13.94 7.612e-09 47-92 BL00203 13.94 7.704e-09 101-146 BL00203 13.94 8.531e-09 75-120 BL00203 13.94 8.714e-09 95-140 BL00203 13.94 9.541e-09 25-70
1090	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 5.613e-09 28-55
1090	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 3.069e-12 26-71 BL00203 13.94 6.266e-10 30-75 BL00203 13.94 4.398e-09 21-66 BL00203 13.94 8.071e-09 25-70
1090	BL00269	Mammalian defensins proteins.	BL00269C 16.52 9.289e-09 32-60
1091	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 8.676e-10 121-163 BL00243I 31.77 3.915e-09 124-166 BL00243I 31.77 5.690e-09 30-72
1091	BL01208	VWFC domain proteins.	BL01208B 15.83 5.865e-09 149-163
1091	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 3.670e-11 66-111 BL00203 13.94 4.659e-11 40-85 BL00203 13.94 7.429e-11 70-115 BL00203 13.94 1.862e-10 105-150 BL00203 13.94 2.723e-10 21-66 BL00203 13.94 2.723e-10 61-106 BL00203 13.94 2.915e-10 126-171 BL00203 13.94 4.064e-10 22-67 BL00203 13.94 6.457e-10 26-71 BL00203 13.94 7.032e-10 149-194 BL00203 13.94 7.319e-10 95-140 BL00203 13.94 1.735e-09 140-185 BL00203 13.94 1.827e-09 115-160 BL00203 13.94 1.918e-09 80-125 BL00203 13.94 3.020e-09 100-145 BL00203 13.94 3.204e-09 65-110 BL00203 13.94 4.306e-09 120-165 BL00203 13.94 5.041e-09 47-92 BL00203 13.94 5.500e-09 116-161 BL00203 13.94 6.694e-09 135-180 BL00203 13.94 6.969e-09 17-62 BL00203 13.94 7.429e-09 71-116 BL00203 13.94 7.704e-09 136-181 BL00203 13.94 8.163e-09 85-130 BL00203 13.94 8.714e-09 130-175
1091	PD02283	PROTEIN SPORULATION REPEAT PRECU.	PD02283C 17.54 5.613e-09 24-51 PD02283C 17.54 5.613e-09 68-95 PD02283C 17.54 7.188e-09 128-155 PD02283C 17.54 7.750e-09 138-165 PD02283C 17.54 8.875e-09 123-150
1091	BL00269	Mammalian defensins proteins.	BL00269C 16.52 9.289e-09 28-56 BL00269C 16.52 9.289e-09 72-100
1091	BL00799	Granulins proteins.	BL00799D 12.41 7.661e-09 49-95 BL00799G 9.41 1.000e-08 39-79

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1094	PR00248	METABOTROPIC GLUTAMATE GPCR SIGNATURE	PR00248A 9.91 7.522e-09 24-45
1094	PR00354	7FE FERREDOXIN SIGNATURE	PR00354C 5.72 8.157e-09 258-275
1096	PR00356	TYPE II ANTIFREEZE PROTEIN SIGNATURE	PR00356G 10.80 9.862e-11 193-206
1096	BL00615	C-type lectin domain proteins.	BL00615B 12.25 2.731e-09 193-206 BL00615A 16.68 9.400e-09 94-111
1097	PR00356	TYPE II ANTIFREEZE PROTEIN SIGNATURE	PR00356G 10.80 7.658e-09 193-206
1097	BL00615	C-type lectin domain proteins.	BL00615A 16.68 9.400e-09 94-111
1098	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 6.870e-24 59-80 PR00245C 7.84 2.421e-19 238-253 PR00245E 12.40 8.714e-16 291-305 PR00245D 10.47 6.786e-13 274-285 PR00245B 10.38 6.906e-13 177-191
1098	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 8.839e-15 90-129 BL00237D 11.23 2.364e-09 282-298
1098	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237B 13.50 1.750e-09 59-80 PR00237C 15.69 4.600e-09 104-126 PR00237A 11.48 5.065e-09 26-50 PR00237G 19.63 5.605e-09 272-298
1098	PR00023	ZONA PELLUCIDA SPERM-BINDING PROTEIN SIGNATURE	PR00023E 22.27 9.813e-09 128-145
1099	DM00191	w SPAC8A4.04C RESISTANCE SPAC8A4.05C DAUNORUBICIN.	DM00191D 13.94 9.083e-10 163-201
1099	PR00346	TISSUE FACTOR SIGNATURE	PR00346H 10.74 8.179e-09 542-565
1099	BL00022	EGF-like domain proteins.	BL00022B 7.54 1.000e-08 306-312
1100	DM00372	CARCINOEMBRYONIC ANTIGEN PRECURSOR AMINO-TERMINAL DOMAIN.	DM00372B 20.31 8.920e-15 363-407 DM00372B 20.31 3.329e-12 68-112
1101	BL01208	VWFC domain proteins.	BL01208B 15.83 3.250e-10 1436-1450
1101	PR00457	ANIMAL HAEM PEROXIDASE SIGNATURE	PR00457E 20.67 3.118e-22 997-1023 PR00457D 16.81 4.194e-21 972-992 PR00457C 19.25 1.675e-13 954-972 PR00457H 15.90 5.680e-13 1248-1262 PR00457F 13.69 4.750e-12 1050-1060 PR00457G 17.45 8.615e-12 1177-1197 PR00457B 13.29 3.411e-10 802-817
1101	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 1.000e-09 349-372
1101	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 7.677e-09 328-367
1101	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 8.920e-09 73-86
1102	BL01208	VWFC domain proteins.	BL01208B 15.83 3.250e-10 1412-1426
1102	PR00457	ANIMAL HAEM PEROXIDASE SIGNATURE	PR00457E 20.67 3.118e-22 973-999 PR00457D 16.81 4.194e-21 948-968 PR00457C 19.25 1.675e-13 930-948 PR00457H 15.90 5.680e-13 1224-1238 PR00457F 13.69 4.750e-12 1026-1036 PR00457G 17.45 8.615e-12 1153-1173 PR00457B 13.29 3.411e-10 778-793
1102	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 1.000e-09 325-348

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1102	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 7.480e-09 73-86
1102	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 7.677e-09 304-343
1103	BL00815	Alpha-isopropylmalate and homocitrate synthases proteins.	BL00815C 21.36 3.118e-09 786-814
1107	PD02059	CORE POLYPROTEIN PROTEIN GAG CONTAINS: P.	PD02059B 24.48 8.352e-09 682-716
1113	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 6.885e-12 311-341
1113	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 7.750e-09 311-329
1117	PD01652	RECEPTOR CELL NK GLYCOPROTEIN IMMUNOGLOB.	PD01652B 8.50 4.021e-09 99-150 PD01652B 8.50 5.050e-09 2-53 PD01652A 15.35 7.769e-09 12-47
1120	BL50002	Src homology 3 (SH3) domain proteins profile.	BL50002A 14.19 1.750e-12 1026-1044
1120	PR00452	SH3 DOMAIN SIGNATURE	PR00452B 11.65 4.115e-11 1036-1051
1120	PF00023	Ank repeat proteins.	PF00023B 14.20 3.000e-10 954-963 PF00023A 16.03 2.286e-09 925-940
1120	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 8.000e-11 951-963 PD00078B 13.14 4.522e-09 918-930
1120	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 8.024e-16 925-979 PF00791C 20.98 4.971e-09 939-977
1120	PR00499	NEUTROPHIL CYTOSOL FACTOR 2 SIGNATURE	PR00499D 10.18 6.965e-09 1024-1044
1122	PF00992	Troponin.	PF00992A 16.67 8.461e-09 245-279
1124	PF00023	Ank repeat proteins.	PF00023A 16.03 7.000e-11 69-84 PF00023B 14.20 2.636e-09 131-140
1124	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 6.087e-09 128-140
1124	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 2.569e-09 135-189 PF00791B 28.49 9.835e-09 69-123
1125	PF00023	Ank repeat proteins.	PF00023A 16.03 7.000e-11 69-84 PF00023B 14.20 2.636e-09 131-140
1125	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 6.087e-09 128-140
1125	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 2.569e-09 135-189 PF00791B 28.49 9.835e-09 69-123
1128	PR00248	METABOTROPIC GLUTAMATE GPCR SIGNATURE	PR00248G 12.67 2.688e-09 53-77
1129	DM00516	186 DISCOIDIN I N-TERMINAL.	DM00516 30.53 8.606e-13 131-175
1130	DM00516	186 DISCOIDIN I N-TERMINAL.	DM00516 30.53 8.606e-13 131-175
1130	DM01077	SEX HORMONE-BINDING GLOBULIN.	DM01077A 16.30 3.143e-11 386-432
1132	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 4.930e-09 87-129
1133	BL00107	Protein kinases ATP-binding region proteins.	BL00107B 13.31 5.909e-13 195-210
1133	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109D 17.04 7.609e-09 196-218 PR00109B 12.27 9.297e-09 126-144
1135	PR00402	TEC/BTK DOMAIN SIGNATURE	PR00402A 16.09 2.950e-10 664-683
1135	BL00509	Ras GTPase-activating proteins.	BL00509B 10.28 9.800e-09 502-512
1137	PR00907	THROMBOMODULIN SIGNATURE	PR00907B 11.29 3.959e-11 168-184
1137	DM00215	PROLINE-RICH PROTEIN 3.	DM00215 19.43 3.893e-10 333-365

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			DM00215 19.43 4.054e-10 328-360 DM00215 19.43 8.232e-10 332-364
1137	BL01187	Calcium-binding EGF-like domain proteins pattern proteins.	BL01187B 12.04 2.957e-13 134-149 BL01187B 12.04 3.739e-13 261-276 BL01187B 12.04 2.333e-12 216-231 BL01187A 9.98 3.250e-09 197-208
1137	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 3.288e-09 348-362 PR00049D 0.00 3.288e-09 350-364
1137	BL01177	Anaphylatoxin domain proteins.	BL01177C 17.39 4.714e-09 128-146
1137	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 5.855e-09 63-88
1137	PF00094	von Willebrand factor type D domain proteins.	PF00094A 11.09 9.022e-09 163-172
1137	BL00022	EGF-like domain proteins.	BL00022B 7.54 9.100e-09 75-81
1137	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 9.357e-09 348-360
1143	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245C 7.84 5.355e-17 121-136 PR00245B 10.38 3.919e-12 60-74 PR00245E 12.40 1.000e-10 174-188
1143	BL00237	G-protein coupled receptors proteins.	BL00237D 11.23 2.091e-09 165-181
1143	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 8.714e-11 155-181 PR00237E 13.03 9.735e-09 82-105
1144	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245C 7.84 5.355e-17 235-250 PR00245A 18.03 8.615e-15 58-79 PR00245B 10.38 3.919e-12 174-188 PR00245E 12.40 1.000e-10 288-302
1144	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 1.581e-15 89-128 BL00237D 11.23 2.091e-09 279-295
1144	PR00896	VASOPRESSIN RECEPTOR SIGNATURE	PR00896B 9.01 8.962e-09 54-65
1144	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 8.714e-11 269-295 PR00237C 15.69 3.829e-10 103-125 PR00237E 13.03 9.735e-09 196-219
1146	BL00914	Syntaxin / epimorphin family proteins.	BL00914 24.91 6.172e-09 168-217
1147	PR00264	INTERLEUKIN-1 SIGNATURE	PR00264B 20.98 8.453e-11 56-82 PR00264C 17.77 1.851e-10 96-124
1148	BL00226	Intermediate filaments proteins.	BL00226B 23.86 5.050e-24 96-143 BL00226D 19.10 8.200e-18 262-308 BL00226C 13.23 5.610e-14 161-191 BL00226A 12.77 5.065e-13 380-394
1151	BL00226	Intermediate filaments proteins.	BL00226D 19.10 5.500e-38 367-413 BL00226C 13.23 4.130e-23 266-296 BL00226A 12.77 9.129e-13 131-145 BL00226B 23.86 1.338e-10 183-230
1152	PR00138	MATRIXIN SIGNATURE	PR00138A 15.14 7.136e-16 86-99 PR00138B 15.82 3.824e-11 131-146
1152	BL00546	Matrixins cysteine switch.	BL00546A 19.62 7.667e-26 66-95 BL00546E 10.23 3.475e-19 231-251 BL00546B 20.11 7.720e-19 155-198 BL00546F 12.40 6.400e-13 268-280 BL00546G 16.84 9.449e-11 288-307
1152	BL00024	Hemopexin domain proteins.	BL00024B 21.53 3.143e-23 105-138

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00024C 22.98 8.320e-20 154-202 BL00024F 11.30 2.184e-18 231-251 BL00024G 13.31 6.192e-13 268-280 BL00024A 11.49 9.100e-13 86-96 BL00024H 11.35 8.154e-10 335-346
1153	PR00138	MATRIXIN SIGNATURE	PR00138A 15.14 7.136e-16 86-99 PR00138B 15.82 3.824e-11 131-146
1153	BL00546	Matrixins cysteine switch.	BL00546A 19.62 7.667e-26 66-95 BL00546E 10.23 3.475e-19 231-251 BL00546B 20.11 7.720e-19 155-198 BL00546F 12.40 6.400e-13 268-280 BL00546G 16.84 9.449e-11 288-307
1153	BL00024	Hemopexin domain proteins.	BL00024B 21.53 3.143e-23 105-138 BL00024C 22.98 8.320e-20 154-202 BL00024F 11.30 2.184e-18 231-251 BL00024G 13.31 6.192e-13 268-280 BL00024A 11.49 9.100e-13 86-96 BL00024H 11.35 8.154e-10 335-346
1154	PR00049	WILM'S TUMOUR PROTEIN SIGNATURE	PR00049D 0.00 2.068e-09 10-24
1155	BL00400	LBP / BPI / CETP family proteins.	BL00400C 24.53 6.029e-17 210-253 BL00400D 23.26 2.080e-14 274-310 BL00400A 21.59 1.600e-10 27-58
1156	PD02448	TRANSCRIPTION PROTEIN DNA-BINDIN.	PD02448A 9.37 1.700e-19 90-128 PD02448B 10.17 2.311e-17 129-176
1156	BL00415	Synapsins proteins.	BL00415O 3.44 7.395e-09 22-59
1159	BL00347	Poly(ADP-ribose) polymerase zinc finger domain proteins.	BL00347A 12.35 9.795e-15 93-135
1159	BL00697	ATP-dependent DNA ligase AMP-binding site proteins.	BL00697D 18.99 1.346e-23 591-617 BL00697A 21.27 2.929e-19 471-499 BL00697B 13.40 4.774e-14 506-517
1160	BL00284	Serpins proteins.	BL00284C 28.56 7.600e-25 203-244 BL00284E 19.15 4.375e-23 401-425 BL00284D 16.34 5.286e-21 317-343 BL00284A 15.64 6.192e-17 27-50 BL00284B 17.99 4.414e-13 174-194
1166	BL01121	Caspase family histidine proteins.	BL01121A 9.11 5.500e-13 7-17
1166	PR00376	INTERLEUKIN-1B CONVERTING ENZYME SIGNATURE	PR00376A 14.23 7.980e-11 5-18
1167	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870D 15.74 7.000e-10 79-113
1167	PD01652	RECEPTOR CELL NK GLYCOPROTEIN IMMUNOGLOB.	PD01652B 8.50 3.143e-29 209-260 PD01652B 8.50 5.457e-18 107-158 PD01652A 15.35 6.438e-14 117-152 PD01652A 15.35 3.732e-10 24-59 PD01652B 8.50 7.448e-10 14-65 PD01652A 15.35 4.231e-09 219-254
1169	BL00615	C-type lectin domain proteins.	BL00615A 16.68 7.231e-10 125-142
1171	PR00308	TYPE I ANTIFREEZE PROTEIN SIGNATURE	PR00308A 5.90 9.156e-13 158-172 PR00308C 3.83 6.640e-12 161-170 PR00308B 4.28 1.806e-10 161-172 PR00308A 5.90 4.873e-10 162-176

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00308C 3.83 8.062e-10 165-174
1171	PR00456	RIBOSOMAL PROTEIN P2 SIGNATURE	PR00456E 3.06 5.671e-09 163-177
1171	BL00678	Trp-Asp (WD) repeat proteins proteins.	BL00678 9.67 2.800e-10 429-439 BL00678 9.67 5.263e-09 480-490 BL00678 9.67 6.211e-09 249-259
1171	PR00833	POLLEN ALLERGEN POA PI SIGNATURE	PR00833H 2.30 7.750e-10 164-178 PR00833H 2.30 7.923e-09 161-175
1171	PR00320	G-PROTEIN BETA WD-40 REPEAT SIGNATURE	PR00320A 16.74 4.000e-13 427-441 PR00320B 12.19 8.269e-12 478-492 PR00320A 16.74 5.966e-11 478-492 PR00320C 13.01 6.478e-11 478-492 PR00320C 13.01 9.217e-11 427-441 PR00320A 16.74 9.690e-11 247-261 PR00320B 12.19 3.057e-10 247-261 PR00320C 13.01 6.040e-10 247-261 PR00320B 12.19 6.657e-10 427-441 PR00320B 12.19 1.450e-09 520-534 PR00320C 13.01 2.500e-09 303-317 PR00320A 16.74 4.732e-09 520-534 PR00320A 16.74 6.488e-09 344-358 PR00320C 13.01 1.000e-08 344-358
1172	PD01652	RECEPTOR CELL NK GLYCOPROTEIN IMMUNOGLOB.	PD01652A 15.35 6.625e-10 24-60 PD01652B 8.50 1.836e-09 14-66 PD01652B 8.50 4.021e-09 111-163
1173	PD01652	RECEPTOR CELL NK GLYCOPROTEIN IMMUNOGLOB.	PD01652A 15.35 6.625e-10 24-60 PD01652B 8.50 1.836e-09 14-66 PD01652B 8.50 4.021e-09 111-163
1183	PD02876	DECARBOXYLASE PHOSPHATIDYLSERINE.	PD02876C 8.80 2.723e-13 316-328 PD02876D 12.13 2.588e-12 427-443
1184	BL01289	TSC-22 / dip / bun family proteins.	BL01289A 12.18 8.200e-33 124-150 BL01289B 10.45 8.071e-30 151-180
1184	DM00475	w LOW TRANSPOSASE SAPA 12K.	DM00475B 12.12 5.891e-10 145-164
1187	PR00901	PHEROMONE B ALPHA-1 RECEPTOR SIGNATURE	PR00901H 14.99 4.706e-09 56-66
1188	BL00708	Prolyl endopeptidase family serine proteins.	BL00708B 24.91 7.197e-12 734-764
1188	PF00930	Dipeptidyl peptidase IV (DPP IV) N-terminal region.	PF00930I 15.96 6.373e-17 776-803 PF00930H 20.16 2.482e-13 697-739 PF00930J 8.78 1.000e-11 828-848 PF00930G 21.30 9.613e-09 657-694
1189	BL00708	Prolyl endopeptidase family serine proteins.	BL00708B 24.91 7.197e-12 734-764
1189	PF00930	Dipeptidyl peptidase IV (DPP IV) N-terminal region.	PF00930H 20.16 2.482e-13 697-739 PF00930J 8.78 1.000e-11 790-810 PF00930G 21.30 9.613e-09 657-694
1190	BL00708	Prolyl endopeptidase family serine proteins.	BL00708B 24.91 7.197e-12 721-751
1190	PF00930	Dipeptidyl peptidase IV (DPP IV) N-terminal region.	PF00930I 15.96 6.373e-17 763-790 PF00930H 20.16 2.482e-13 684-726 PF00930J 8.78 1.000e-11 815-835 PF00930G 21.30 9.613e-09 644-681
1193	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 6.612e-15 153-207 PF00791B 28.49 7.955e-14 186-240 PF00791B 28.49 3.653e-12 436-490

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PF00791B 28.49 9.337e-12 54-108 PF00791B 28.49 4.273e-11 319-373 PF00791B 28.49 7.818e-11 252-306 PF00791B 28.49 1.524e-10 219-273 PF00791B 28.49 2.398e-10 120-174 PF00791C 20.98 3.559e-09 200-238 PF00791C 20.98 5.235e-09 333-371 PF00791C 20.98 5.235e-09 544-582 PF00791B 28.49 6.202e-09 352-406 PF00791B 28.49 7.028e-09 598-652 PF00791C 20.98 7.265e-09 101-139 PF00791B 28.49 8.679e-09 530-584 PF00791B 28.49 1.000e-08 87-141
1193	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 4.600e-12 345-357 PD00078B 13.14 2.000e-11 462-474 PD00078B 13.14 3.500e-11 796-808 PD00078B 13.14 8.500e-11 863-875 PD00078B 13.14 4.600e-10 495-507 PD00078B 13.14 5.950e-10 760-772 PD00078B 13.14 4.522e-09 212-224 PD00078B 13.14 6.087e-09 278-290 PD00078B 13.14 1.000e-08 146-158 PD00078B 13.14 1.000e-08 245-257
1193	PF00023	Ank repeat proteins.	PF00023A 16.03 2.500e-12 186-201 PF00023B 14.20 5.154e-11 465-474 PF00023B 14.20 5.154e-11 763-772 PF00023A 16.03 6.571e-11 153-168 PF00023A 16.03 1.750e-10 54-69 PF00023B 14.20 8.000e-10 866-875 PF00023B 14.20 1.409e-09 348-357 PF00023B 14.20 2.636e-09 281-290 PF00023A 16.03 3.250e-09 219-234 PF00023B 14.20 3.455e-09 498-507 PF00023B 14.20 3.864e-09 799-808 PF00023A 16.03 4.536e-09 252-267 PF00023B 14.20 5.500e-09 248-257 PF00023A 16.03 6.464e-09 598-613 PF00023B 14.20 7.955e-09 432-441 PF00023A 16.03 8.071e-09 631-646 PF00023A 16.03 8.071e-09 767-782 PF00023A 16.03 1.000e-08 701-716
1194	PR00834	HTRA/DEGQ PROTEASE FAMILY SIGNATURE	PR00834C 15.43 6.226e-20 253-277 PR00834D 12.14 4.316e-17 291-308 PR00834B 10.09 7.188e-14 212-232 PR00834E 13.63 1.000e-12 313-330 PR00834A 9.80 5.737e-12 191-203 PR00834F 10.91 1.730e-09 374-386
1195	PR00555	ADENOSINE A3 RECEPTOR SIGNATURE	PR00555E 11.12 5.629e-20 105-122 PR00555F 11.18 6.114e-20 152-169 PR00555D 10.11 4.717e-18 60-76
1195	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 8.560e-15 119-145 PR00237F 13.57 3.520e-13 83-107

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00237E 13.03 4.960e-12 33-56
1195	PR00424	ADENOSINE RECEPTOR SIGNATURE	PR00424D 14.32 9.400e-23 21-40 PR00424E 15.73 6.211e-14 74-87 PR00424F 8.50 9.156e-12 119-129
1195	BL00237	G-protein coupled receptors proteins.	BL00237C 13.19 3.864e-15 78-104 BL00237D 11.23 1.346e-11 129-145
1197	BL00237	G-protein coupled receptors proteins.	BL00237A 27.68 3.455e-14 95-134
1197	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237C 15.69 1.257e-10 109-131 PR00237E 13.03 9.100e-10 204-227
1197	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 9.581e-18 64-85 PR00245C 7.84 4.780e-13 243-258 PR00245E 12.40 6.741e-09 296-310 PR00245B 10.38 8.163e-09 182-196
1197	PR00534	MELANOCORTIN RECEPTOR FAMILY SIGNATURE	PR00534A 11.49 9.229e-09 56-68
1198	PR00505	D12 CLASS N6 ADENINE-SPECIFIC DNA METHYLTRANSFERASE SIGNATURE	PR00505A 14.15 4.857e-13 30-46 PR00505B 11.49 1.621e-12 51-65
1199	PR00179	LIPOCALIN SIGNATURE	PR00179B 9.56 2.071e-09 111-123 PR00179C 19.02 9.455e-09 138-153
1200	PF00152	tRNA synthetases class II.	PF00152D 21.30 8.364e-28 431-469 PF00152C 28.03 9.250e-21 220-256 PF00152B 15.67 2.658e-13 159-183 PF00152A 19.68 5.714e-11 44-66
1202	BL00504	Fumarate reductase / succinate dehydrogenase FAD-binding site proteins.	BL00504D 10.43 5.390e-17 31-48
1203	BL00720	Guanine-nucleotide dissociation stimulators CDC25 family sign.	BL00720B 16.57 5.065e-17 309-332
1204	PF00013	KH domain proteins family of RNA binding proteins.	PF00013 5.78 4.150e-09 112-123
1206	DM00893	YRUVATE DEHYDROGENASE (LIPOAMIDE) BETA CHAIN.	DM00893A 19.01 1.000e-40 47-93 DM00893E 29.52 1.000e-40 234-287 DM00893C 20.28 2.452e-40 143-184 DM00893B 27.53 3.483e-31 105-142 DM00893D 23.36 1.545e-26 197-230 DM00893F 21.02 6.897e-21 292-316
1207	PR00312	CALSEQUESTRIN SIGNATURE	PR00312E 8.32 3.423e-36 163-192 PR00312I 15.78 5.286e-35 326-354 PR00312F 15.06 5.865e-35 193-222 PR00312H 13.31 8.313e-35 257-284 PR00312J 13.73 5.688e-34 357-385 PR00312D 9.43 2.636e-33 122-151 PR00312C 15.14 8.839e-33 86-115 PR00312B 15.08 8.941e-33 56-85 PR00312G 11.11 6.657e-32 224-251 PR00312A 11.70 6.914e-27 29-52
1207	BL00863	Calsequestrin proteins.	BL00863G 12.17 1.000e-40 192-233 BL00863H 14.03 1.000e-40 240-276 BL00863J 10.84 1.000e-40 304-341 BL00863A 15.14 7.387e-40 28-64 BL00863B 12.89 4.300e-32 65-92 BL00863F 11.27 3.172e-31 161-187

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00863I 10.28 6.786e-31 277-303 BL00863E 8.49 1.462e-28 135-160 BL00863C 13.93 7.387e-24 93-114 BL00863D 11.58 5.629e-19 115-132
1209	BL00781	Phosphoenolpyruvate carboxylase proteins 1.	BL00781C 12.88 7.031e-09 233-287
1209	PR00985	LEUCYL-TRNA SYNTHETASE SIGNATURE	PR00985A 12.10 7.716e-09 515-532
1209	PR00563	BETA-3 ADRENERGIC RECEPTOR SIGNATURE	PR00563E 7.48 8.768e-09 782-800
1210	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290A 20.89 1.818e-11 158-180
1213	BL00232	Cadherins extracellular repeat proteins domain proteins.	BL00232B 32.79 2.125e-26 227-274 BL00232B 32.79 8.521e-15 440-487 BL00232B 32.79 1.346e-13 118-165 BL00232B 32.79 5.500e-13 335-382 BL00232C 10.65 7.923e-10 333-350 BL00232C 10.65 9.308e-10 438-455 BL00232C 10.65 9.827e-10 225-242
1213	PR00205	CADHERIN SIGNATURE	PR00205B 11.39 3.945e-10 438-455 PR00205B 11.39 2.220e-09 333-350 PR00205B 11.39 9.542e-09 548-565
1214	PR00626	CALRETICULIN SIGNATURE	PR00626D 8.30 8.071e-30 242-264 PR00626E 11.30 7.632e-24 280-299 PR00626B 14.12 2.200e-20 126-142 PR00626E 11.30 3.676e-19 266-285 PR00626A 14.35 1.500e-18 100-118 PR00626C 9.70 9.100e-18 215-228 PR00626C 9.70 7.882e-14 232-245 PR00626D 8.30 8.017e-13 256-278 PR00626D 8.30 6.520e-09 208-230
1214	BL00803	Calreticulin family proteins.	BL00803G 14.33 1.000e-40 258-302 BL00803F 10.95 2.000e-37 225-255 BL00803E 16.55 2.588e-31 166-196 BL00803C 11.13 6.063e-26 91-113 BL00803F 10.95 7.268e-22 208-238 BL00803G 14.33 1.127e-19 244-288 BL00803B 17.08 8.714e-18 63-81 BL00803D 16.08 1.000e-15 128-138 BL00803G 14.33 3.962e-15 272-316 BL00803A 14.83 2.688e-14 35-48 BL00803F 10.95 2.179e-11 191-221 BL00803F 10.95 9.516e-09 242-272
1215	PF00711	Beta defensins.	PF00711 15.76 7.915e-11 45-77
1215	PD00866	GLYCOPROTEIN PROTEIN SPIKE E2 PRECURSOR PEPLIMER.	PD00866L 3.73 7.709e-10 59-68
1215	PR00858	CRUSTACEAN METALLOTHIONEIN SIGNATURE	PR00858B 5.93 1.479e-09 40-58
1215	BL00317	WAP-type 'four-disulfide core' domain proteins.	BL00317B 14.58 2.216e-09 48-69
1215	BL00264	Neurohypophysial hormones proteins.	BL00264 8.98 5.642e-09 79-105
1215	DM01724	kw ALLERGEN POLLEN CIM1 HOL-LI.	DM01724 8.14 7.968e-12 16-35

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			DM01724 8.14 1.409e-11 20-39 DM01724 8.14 1.507e-10 4-23 DM01724 8.14 6.684e-09 12-31
1215	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243I 31.77 2.000e-11 42-84 BL00243I 31.77 1.265e-10 54-96 BL00243I 31.77 1.254e-09 45-87 BL00243I 31.77 8.225e-09 58-100
1215	BL00203	Vertebrate metallothioneins proteins.	BL00203 13.94 2.862e-12 32-77 BL00203 13.94 3.690e-12 39-84 BL00203 13.94 4.758e-11 35-80 BL00203 13.94 3.663e-09 42-87 BL00203 13.94 5.592e-09 50-95 BL00203 13.94 6.235e-09 36-81 BL00203 13.94 6.786e-09 40-85 BL00203 13.94 9.357e-09 60-105
1218	PR00946	MERCURY SCAVENGER PROTEIN SIGNATURE	PR00946A 5.58 6.516e-10 6-24
1220	DM01071	OPACITY PROTEIN.	DM01071A 1.92 8.990e-09 5-20
1221	BL00884	Osteopontin proteins.	BL00884C 22.45 1.000e-40 119-160 BL00884B 12.47 4.673e-33 24-67 BL00884A 11.35 8.615e-32 1-30 BL00884D 8.79 4.857e-19 248-264
1221	PR00216	OSTEOPONTIN SIGNATURE	PR00216A 10.94 5.000e-35 2-31 PR00216C 9.63 1.391e-32 41-66 PR00216G 12.39 9.550e-31 231-256 PR00216F 11.79 3.700e-23 152-170 PR00216E 8.44 3.250e-19 120-134 PR00216D 2.74 1.200e-18 88-102 PR00216D 2.74 2.209e-12 82-96
1222	BL00284	Serpins proteins.	BL00284C 28.56 6.538e-29 225-266 BL00284A 15.64 3.739e-18 107-130 BL00284D 16.34 3.793e-17 332-358 BL00284E 19.15 2.909e-15 419-443
1223	PD02327	GLYCOPROTEIN ANTIGEN PRECURSOR IMMUNOGLO.	PD02327B 19.84 8.941e-23 143-164 PD02327A 8.89 1.000e-13 115-126 PD02327C 15.47 5.500e-13 209-223
1225	PR00418	DNA TOPOISOMERASE II SIGNATURE	PR00418F 12.01 3.813e-20 470-486 PR00418G 14.68 7.000e-19 488-505 PR00418C 10.02 8.200e-18 100-114 PR00418I 16.64 4.682e-17 550-566 PR00418A 12.34 3.739e-16 20-35 PR00418B 12.52 6.571e-15 57-70 PR00418E 15.56 7.300e-15 397-411 PR00418D 14.93 7.000e-14 252-265 PR00418H 13.54 2.385e-12 508-520
1225	BL00177	DNA topoisomerase II proteins.	BL00177H 21.42 3.647e-39 471-506 BL00177G 24.83 4.706e-36 417-455 BL00177B 19.24 1.000e-35 79-114 BL00177I 21.82 2.200e-21 732-757 BL00177F 12.98 2.500e-18 395-412 BL00177D 14.66 9.591e-15 252-265 BL00177E 12.43 7.000e-13 310-321

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			BL00177C 13.16 5.950e-12 155-166
1225	BL01190	Ribosomal protein L36e proteins.	BL01190B 16.17 6.929e-10 1140-1194
1225	PF00521	DNA gyrase/topoisomerase IV, subunit A.	PF00521D 9.77 9.591e-09 788-811
1226	BL00455	Putative AMP-binding domain proteins.	BL00455 13.31 6.684e-13 248-263
1226	PR00154	AMP-BINDING SIGNATURE	PR00154A 8.88 7.375e-10 241-252
1228	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007B 14.16 7.698e-13 116-135 PR00007D 9.64 9.654e-11 193-203 PR00007A 19.33 2.552e-10 89-115 PR00007C 15.60 3.656e-10 163-184
1228	BL01113	C1q domain proteins.	BL01113B 18.26 1.563e-20 95-130 BL01113D 7.47 9.308e-12 195-204 BL01113C 13.18 4.750e-10 163-182
1230	PD00078	REPEAT PROTEIN ANK NUCLEAR ANKYR.	PD00078B 13.14 1.000e-11 378-390 PD00078B 13.14 4.500e-11 495-507 PD00078B 13.14 8.200e-10 897-909 PD00078B 13.14 4.522e-09 528-540
1230	PR00665	OXYTOCIN RECEPTOR SIGNATURE	PR00665E 5.60 5.390e-09 756-769
1230	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 1.890e-13 186-240 PF00791B 28.49 3.368e-12 469-523 PF00791B 28.49 2.273e-11 219-273 PF00791B 28.49 2.922e-10 352-406 PF00791B 28.49 3.534e-10 904-958 PF00791C 20.98 5.361e-10 366-404 PF00791B 28.49 8.427e-10 12-66 PF00791B 28.49 8.951e-10 734-788 PF00791B 28.49 2.156e-09 153-207 PF00791B 28.49 7.028e-09 563-617
1230	PF00023	Ank repeat proteins.	PF00023A 16.03 1.600e-13 219-234 PF00023A 16.03 2.500e-12 252-267 PF00023B 14.20 5.154e-11 498-507 PF00023A 16.03 7.750e-10 631-646 PF00023B 14.20 8.000e-10 900-909 PF00023A 16.03 1.321e-09 186-201 PF00023B 14.20 1.409e-09 381-390 PF00023A 16.03 2.607e-09 698-713 PF00023B 14.20 4.273e-09 465-474 PF00023A 16.03 4.536e-09 1007-1022 PF00023B 14.20 5.500e-09 281-290 PF00023B 14.20 7.545e-09 531-540 PF00023A 16.03 1.000e-08 800-815
1231	BL00400	LBP / BPI / CETP family proteins.	BL00400C 24.53 6.029e-17 210-253 BL00400D 23.26 2.080e-14 274-310 BL00400A 21.59 1.600e-10 27-58
1232	BL00400	LBP / BPI / CETP family proteins.	BL00400C 24.53 6.029e-17 210-253 BL00400D 23.26 2.080e-14 274-310 BL00400A 21.59 1.600e-10 27-58
1233	BL00400	LBP / BPI / CETP family proteins.	BL00400C 24.53 6.029e-17 210-253 BL00400D 23.26 2.080e-14 274-310 BL00400A 21.59 1.600e-10 27-58
1237	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 9.809e-09 132-155
1247	BL01248	Laminin-type EGF-like (LE) domain proteins.	BL01248 11.02 1.340e-09 289-301

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1247	PR00764	COMPLEMENT C9 SIGNATURE	PR00764F 16.89 6.610e-09 237-257
1247	BL00812	Glycosyl hydrolases family 8 proteins.	BL00812B 13.49 6.667e-09 917-931
1247	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011B 13.08 9.386e-17 767-785 PR00011B 13.08 8.875e-16 289-307 PR00011D 14.03 5.800e-15 550-568 PR00011D 14.03 8.000e-15 767-785 PR00011D 14.03 3.388e-14 289-307 PR00011B 13.08 7.833e-14 160-178 PR00011B 13.08 9.000e-14 550-568 PR00011A 14.06 9.345e-14 289-307 PR00011B 13.08 5.119e-13 203-221 PR00011B 13.08 5.576e-13 421-439 PR00011D 14.03 6.943e-13 421-439 PR00011B 13.08 7.102e-13 638-656 PR00011A 14.06 9.237e-13 203-221 PR00011B 13.08 9.542e-13 378-396 PR00011D 14.03 9.830e-13 638-656 PR00011D 14.03 3.211e-12 378-396 PR00011B 13.08 4.339e-12 810-828 PR00011A 14.06 6.516e-12 378-396 PR00011D 14.03 6.842e-12 810-828 PR00011D 14.03 7.158e-12 160-178 PR00011A 14.06 8.548e-12 421-439 PR00011A 14.06 1.554e-11 550-568 PR00011D 14.03 2.770e-11 593-611 PR00011D 14.03 3.213e-11 507-525 PR00011D 14.03 3.361e-11 203-221 PR00011B 13.08 4.877e-11 246-264 PR00011B 13.08 6.400e-11 332-350 PR00011B 13.08 6.815e-11 593-611 PR00011D 14.03 7.049e-11 332-350 PR00011B 13.08 8.062e-11 724-742 PR00011B 13.08 2.174e-10 507-525 PR00011D 14.03 2.523e-10 464-482 PR00011A 14.06 3.348e-10 767-785 PR00011D 14.03 4.462e-10 724-742 PR00011A 14.06 5.304e-10 810-828 PR00011A 14.06 8.304e-10 638-656 PR00011D 14.03 8.892e-10 246-264 PR00011D 14.03 1.913e-09 681-699 PR00011B 13.08 2.356e-09 464-482 PR00011A 14.06 2.726e-09 160-178 PR00011A 14.06 2.849e-09 246-264 PR00011B 13.08 5.685e-09 681-699 PR00011A 14.06 5.808e-09 681-699 PR00011A 14.06 6.055e-09 724-742 PR00011A 14.06 6.425e-09 464-482 PR00011A 14.06 6.671e-09 507-525
1247	DM00758	AGRIN.	DM00758 13.12 7.485e-09 197-212 DM00758 13.12 8.412e-09 240-255
1247	PR00173	GLUTAMATE-ASPARTATE SYMPORTER SIGNATURE	PR00173F 10.44 8.820e-09 859-878

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1247	BL00022	EGF-like domain proteins.	BL00022B 7.54 3.250e-10 210-216 BL00022A 7.48 9.000e-09 283-289
1247	BL00243	Integrins beta chain cysteine-rich domain proteins.	BL00243H 17.53 4.671e-09 284-309 BL00243H 17.53 7.750e-09 327-352 BL00243H 17.53 8.816e-09 198-223 BL00243H 17.53 9.053e-09 241-266
1254	BL00247	HBGF/FGF family proteins.	BL00247B 31.59 3.077e-35 82-128 BL00247C 21.54 8.333e-22 137-164
1254	PR00262	IL1/HBGF FAMILY SIGNATURE	PR00262A 28.26 8.588e-11 77-104
1254	PR00263	HEPARIN BINDING GROWTH FACTOR FAMILY SIGNATURE	PR00263D 12.89 5.078e-11 106-125 PR00263C 9.90 7.188e-10 90-102
1260	PR00345	STATHMIN FAMILY SIGNATURE	PR00345B 7.12 1.371e-11 207-235
1260	BL00563	Stathmin family proteins.	BL00563B 6.08 6.021e-11 213-239
1260	PF00780	Domain found in NIK1-like kinases, mouse citron and yeast ROM.	PF00780A 10.77 7.857e-10 68-76
1260	BL00326	Tropomyosins proteins.	BL00326B 7.68 1.235e-09 161-209
1260	PR00194	TROPOMYOSIN SIGNATURE	PR00194C 6.38 9.703e-09 120-148
1261	BL00284	Serpins proteins.	BL00284C 28.56 7.000e-17 212-253 BL00284D 16.34 1.692e-13 324-350 BL00284A 15.64 1.200e-11 49-72
1262	BL00873	Sodium:alanine symporter family proteins.	BL00873B 20.93 9.029e-10 2-53
1263	BL01020	SAR1 family proteins.	BL01020C 15.35 3.506e-20 83-133 BL01020A 11.87 3.821e-19 7-37 BL01020B 11.70 5.393e-15 41-75
1263	PR00328	GTP-BINDING SAR1 PROTEIN SIGNATURE	PR00328B 9.04 2.112e-12 55-79 PR00328A 10.62 4.857e-12 27-50
1265	PR00258	SPERACT RECEPTOR SIGNATURE	PR00258B 9.63 2.800e-14 493-504 PR00258C 9.05 1.257e-12 62-72 PR00258C 9.05 7.171e-12 508-518 PR00258D 14.41 8.500e-12 539-553 PR00258D 14.41 8.875e-12 93-107 PR00258A 11.46 3.418e-10 229-245 PR00258D 14.41 5.034e-10 294-308 PR00258E 13.33 2.500e-09 215-227 PR00258A 11.46 3.000e-09 133-149 PR00258C 9.05 7.000e-09 163-173
1265	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 1.000e-40 478-532 BL00420B 22.67 7.689e-25 233-287 BL00420B 22.67 6.625e-18 32-86 BL00420B 22.67 8.863e-15 133-187 BL00420B 22.67 5.585e-12 361-415 BL00420C 11.90 8.625e-09 216-226 BL00420C 11.90 9.000e-09 563-573
1266	PR00258	SPERACT RECEPTOR SIGNATURE	PR00258B 9.63 2.800e-14 493-504 PR00258C 9.05 1.257e-12 62-72 PR00258C 9.05 7.171e-12 508-518 PR00258D 14.41 8.500e-12 539-553 PR00258D 14.41 8.875e-12 93-107 PR00258A 11.46 3.418e-10 229-245 PR00258D 14.41 5.034e-10 294-308 PR00258E 13.33 2.500e-09 215-227

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00258A 11.46 3.000e-09 133-149 PR00258C 9.05 7.000e-09 163-173
1266	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 1.000e-40 478-532 BL00420B 22.67 7.689e-25 233-287 BL00420B 22.67 6.625e-18 32-86 BL00420B 22.67 8.863e-15 133-187 BL00420B 22.67 5.585e-12 361-415 BL00420C 11.90 8.625e-09 216-226 BL00420C 11.90 9.000e-09 563-573
1272	PR00170	SODIUM CHANNEL SIGNATURE	PR00170E 6.48 8.533e-09 34-63
1273	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 5.500e-21 214-244
1273	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109B 12.27 9.294e-12 214-232
1273	BL00239	Receptor tyrosine kinase class II proteins.	BL00239B 25.15 2.935e-09 149-196
1273	BL00240	Receptor tyrosine kinase class III proteins.	BL00240E 11.56 1.000e-08 200-237
1275	BL00427	Disintegrins proteins.	BL00427 13.93 7.592e-26 460-514
1275	PR00138	MATRIXIN SIGNATURE	PR00138D 16.56 5.101e-11 359-384
1275	BL00142	Neutral zinc metallopeptidases, zinc-binding region proteins.	BL00142 8.38 7.545e-11 359-369
1275	PR00289	DISINTEGRIN SIGNATURE	PR00289A 13.62 2.500e-14 474-493 PR00289B 11.79 4.226e-10 503-515
1275	PR00480	ASTACIN FAMILY SIGNATURE	PR00480B 15.41 8.909e-10 354-372
1275	PR00907	THROMBOMODULIN SIGNATURE	PR00907E 11.70 3.647e-09 672-694
1275	BL00546	Matrixins cysteine switch.	BL00546C 16.41 4.255e-09 353-384
1275	BL00024	Hemopexin domain proteins.	BL00024D 17.28 5.596e-09 353-384
1277	PF00023	Ank repeat proteins.	PF00023A 16.03 1.600e-13 345-360 PF00023B 14.20 6.318e-09 302-311 PF00023A 16.03 6.464e-09 306-321
1278	BL00142	Neutral zinc metallopeptidases, zinc-binding region proteins.	BL00142 8.38 1.857e-09 412-422
1278	PR00756	MEMBRANE ALANYL DIPEPTIDASE (M1) FAMILY SIGNATURE	PR00756A 12.90 5.091e-17 245-260 PR00756D 10.58 8.258e-17 412-427 PR00756B 14.06 7.333e-14 297-312 PR00756E 11.91 3.769e-09 431-443
1279	DM01688	2 POLY-IG RECEPTOR.	DM01688K 17.19 8.640e-11 78-116 DM01688G 16.45 5.680e-09 76-107
1288	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 8.043e-10 164-177 PR00019B 11.36 7.120e-09 136-149
1288	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 7.319e-09 319-342
1290	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019A 11.19 3.400e-12 86-99 PR00019B 11.36 9.357e-12 83-96 PR00019A 11.19 4.333e-09 111-124
1295	BL01113	C1q domain proteins.	BL01113C 13.18 9.617e-13 159-178 BL01113D 7.47 2.174e-11 191-200 BL01113B 18.26 7.658e-11 91-126 BL01113A 17.99 3.106e-10 22-48
1295	PR00007	COMPLEMENT C1Q DOMAIN SIGNATURE	PR00007B 14.16 9.769e-14 112-131 PR00007C 15.60 5.688e-13 159-180 PR00007D 9.64 1.419e-09 189-199 PR00007A 19.33 4.429e-09 86-112
1295	PR00513	5-HYDROXYTRYPTAMINE 1B	PR00513D 11.06 8.085e-09 50-67

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
		RECEPTOR SIGNATURE	
1296	PR00665	OXYTOCIN RECEPTOR SIGNATURE	PR00665D 9.93 9.012e-11 108-124
1296	BL00896	LacY family proton/sugar symporters proteins.	BL00896A 14.92 2.552e-09 300-332
1296	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237F 13.57 8.667e-12 269-293 PR00237G 19.63 7.395e-10 314-340 PR00237A 11.48 8.333e-10 34-58 PR00237B 13.50 4.250e-09 68-89
1296	BL00237	G-protein coupled receptors proteins.	BL00237C 13.19 4.414e-12 264-290 BL00237D 11.23 9.727e-09 324-340
1297	BL00019	Actinin-type actin-binding domain proteins.	BL00019C 14.66 6.250e-28 285-320 BL00019D 15.33 2.309e-15 348-377 BL00019B 13.34 2.976e-13 240-262 BL00019A 12.56 2.286e-12 215-225
1297	PF00435	Spectrin repeat proteins.	PF00435A 32.05 2.000e-14 991-1019 PF00435B 13.41 9.609e-11 1496-1511 PF00435C 20.73 3.571e-09 2006-2025
1297	DM00588	8 kw CHO2 ALPHA ANTIGEN PARAMYOSIN.	DM00588B 9.45 6.870e-09 1259-1268
1297	BL00326	Tropomyosins proteins.	BL00326B 7.68 9.296e-09 2110-2158
1297	BL00226	Intermediate filaments proteins.	BL00226B 23.86 5.605e-09 1734-1781 BL00226B 23.86 9.895e-09 2042-2089
1298	BL00019	Actinin-type actin-binding domain proteins.	BL00019C 14.66 6.250e-28 297-332 BL00019D 15.33 2.309e-15 360-389 BL00019B 13.34 2.976e-13 240-262 BL00019A 12.56 2.286e-12 215-225
1298	PF00435	Spectrin repeat proteins.	PF00435A 32.05 2.000e-14 1003-1031 PF00435B 13.41 9.609e-11 1508-1523 PF00435C 20.73 3.571e-09 2018-2037
1298	DM00588	8 kw CHO2 ALPHA ANTIGEN PARAMYOSIN.	DM00588B 9.45 6.870e-09 1271-1280
1298	BL00326	Tropomyosins proteins.	BL00326B 7.68 9.296e-09 2122-2170
1298	BL00226	Intermediate filaments proteins.	BL00226B 23.86 5.605e-09 1746-1793 BL00226B 23.86 9.895e-09 2054-2101
1304	PR00700	PROTEIN TYROSINE PHOSPHATASE SIGNATURE	PR00700C 13.17 8.535e-09 125-142
1305	PR00700	PROTEIN TYROSINE PHOSPHATASE SIGNATURE	PR00700C 13.17 8.535e-09 240-257
1306	PD02929	ADHESION GLYCOPROTEIN PRECURSOR I.	PD02929A 28.27 4.433e-10 207-260
1306	PR00020	MAM DOMAIN SIGNATURE	PR00020A 18.17 9.211e-10 428-446 PR00020C 13.66 3.340e-09 509-520
1306	BL00740	MAM domain proteins.	BL00740B 19.76 4.682e-10 578-598 BL00740A 13.87 5.588e-09 430-442
1308	BL00072	Acyl-CoA dehydrogenases proteins.	BL00072E 24.12 5.014e-12 724-766 BL00072D 30.08 7.136e-10 635-685
1309	BL00072	Acyl-CoA dehydrogenases proteins.	BL00072E 24.12 5.014e-12 706-748 BL00072D 30.08 7.136e-10 617-667
1311	PR00215	NEUROMODULIN SIGNATURE	PR00215C 13.98 6.779e-10 743-763
1311	BL00412	Neuromodulin (GAP-43) proteins.	BL00412B 10.60 1.681e-09 735-771
1311	PF00992	Troponin.	PF00992A 16.67 9.746e-10 609-643

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PF00992A 16.67 5.145e-09 613-647 PF00992A 16.67 7.395e-09 615-649 PF00992A 16.67 1.000e-08 608-642
1314	PF00632	HECT-domain (ubiquitin-transferase).	PF00632C 20.66 1.000e-29 2270-2301 PF00632B 18.45 2.800e-21 2215-2242
1314	PF00624	Flocculin repeat proteins.	PF00624J 6.21 7.000e-09 1424-1478
1314	BL00412	Neuromodulin (GAP-43) proteins.	BL00412D 16.54 9.022e-10 350-400 BL00412D 16.54 1.551e-09 342-392 BL00412D 16.54 7.429e-09 349-399 BL00412D 16.54 8.531e-09 328-378
1314	DM00191	w SPAC8A4.04C RESISTANCE SPAC8A4.05C DAUNORUBICIN.	DM00191D 13.94 6.635e-09 1410-1448 DM00191D 13.94 9.374e-09 1404-1442
1317	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 5.263e-10 107-116
1321	PR00019	LEUCINE-RICH REPEAT SIGNATURE	PR00019B 11.36 4.000e-11 335-348 PR00019B 11.36 1.450e-10 193-206 PR00019B 11.36 3.250e-10 167-180 PR00019A 11.19 4.130e-10 338-351 PR00019A 11.19 4.522e-10 480-493 PR00019B 11.36 7.300e-10 309-322 PR00019B 11.36 1.720e-09 569-582 PR00019B 11.36 3.880e-09 477-490 PR00019A 11.19 5.667e-09 170-183
1321	DM01551	kw OSTEOINDUCTIVE YOPM MEMBRANE OUTER.	DM01551C 14.62 6.280e-09 568-587 DM01551C 14.62 8.320e-09 355-374
1322	BL00290	Immunoglobulins and major histocompatibility complex proteins.	BL00290B 13.17 9.250e-09 317-334
1324	PD01719	PRECURSOR GLYCOPROTEIN SIGNAL RE.	PD01719A 12.89 1.740e-11 36-63
1328	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 4.696e-38 15-69 BL00420B 22.67 6.949e-36 189-243 BL00420B 22.67 1.300e-35 301-355 BL00420B 22.67 4.358e-30 639-693 BL00420B 22.67 1.863e-26 406-460 BL00420C 11.90 1.360e-13 100-110 BL00420C 11.90 6.797e-11 274-284 BL00420C 11.90 8.322e-11 492-502 BL00420C 11.90 1.545e-10 386-396
1328	PR00258	SPERACT RECEPTOR SIGNATURE	PR00258B 9.63 7.188e-15 654-665 PR00258B 9.63 8.875e-15 30-41 PR00258B 9.63 8.875e-15 204-215 PR00258B 9.63 6.400e-14 316-327 PR00258B 9.63 3.543e-13 421-432 PR00258E 13.33 7.811e-13 99-111 PR00258D 14.41 7.500e-11 468-482 PR00258E 13.33 9.625e-11 273-285 PR00258D 14.41 2.552e-10 700-714 PR00258E 13.33 3.000e-10 491-503 PR00258A 11.46 8.791e-10 635-651 PR00258C 9.05 1.000e-09 45-55 PR00258A 11.46 2.375e-09 185-201 PR00258A 11.46 6.500e-09 11-27 PR00258A 11.46 6.500e-09 297-313

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
			PR00258E 13.33 7.450e-09 385-397 PR00258C 9.05 8.500e-09 436-446 PR00258A 11.46 9.625e-09 402-418
1329	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270A 17.22 7.500e-15 21-60 PD01270B 22.18 6.288e-13 72-108 PD01270C 19.54 7.608e-09 114-142
1333	BL00246	Wnt-1 family proteins.	BL00246D 23.97 1.000e-40 202-254 BL00246E 20.32 8.636e-35 319-364 BL00246B 13.69 6.806e-29 101-135 BL00246C 15.56 9.036e-22 167-191 BL00246A 15.75 6.870e-21 68-87
1335	PR00245	OLFACTORY RECEPTOR SIGNATURE	PR00245A 18.03 7.300e-19 26-47
1337	BL00476	Fatty acid desaturases family 1 proteins.	BL00476C 13.87 1.000e-40 80-132 BL00476E 12.10 1.000e-40 231-283 BL00476D 11.28 2.125e-30 171-221 BL00476B 18.34 4.494e-16 36-79 BL00476F 12.75 6.333e-16 285-329
1337	PR00075	FATTY ACID DESATURASE FAMILY 1 SIGNATURE	PR00075D 11.41 3.538e-33 131-160 PR00075C 10.31 3.813e-20 94-114 PR00075G 8.85 2.047e-19 268-282 PR00075E 12.60 7.585e-16 192-210 PR00075F 16.07 6.952e-15 225-246 PR00075A 16.97 4.429e-14 47-67 PR00075B 12.16 7.047e-11 71-93
1339	PD00301	PROTEIN REPEAT MUSCLE CALCIUM-BL.	PD00301A 10.24 6.400e-09 55-65
1339	BL00422	Granins proteins.	BL00422C 16.18 6.647e-09 44-71 BL00422C 16.18 8.235e-09 45-72
1339	BL00319	Amyloidogenic glycoprotein extracellular domain proteins.	BL00319C 17.12 5.836e-11 48-81 BL00319C 17.12 5.974e-09 47-80 BL00319C 17.12 8.342e-09 44-77 BL00319C 17.12 9.053e-09 45-78
1340	BL00406	Actins proteins.	BL00406C 6.75 4.286e-20 137-191 BL00406B 5.47 8.130e-14 78-132 BL00406D 12.58 3.734e-13 267-321 BL00406A 9.95 1.290e-12 5-39
1340	PR00190	ACTIN SIGNATURE	PR00190F 7.80 4.803e-12 135-154 PR00190C 11.49 1.878e-09 57-79
1341	BL00048	Protamine P1 proteins.	BL00048 6.39 3.588e-09 4-30
1343	BL00790	Receptor tyrosine kinase class V proteins.	BL00790I 20.01 9.520e-11 555-585
1343	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014C 15.44 2.565e-09 544-562
1344	PR00020	MAM DOMAIN SIGNATURE	PR00020A 18.17 5.776e-12 759-777 PR00020C 13.66 6.932e-10 832-843
1344	PD01270	RECEPTOR FC IMMUNOGLOBULIN AFFIN.	PD01270D 24.66 5.378e-09 292-327
1344	BL00740	MAM domain proteins.	BL00740A 13.87 8.313e-12 761-773 BL00740B 19.76 8.500e-09 901-921
1344	PD02080	T-CELL GLYCOPROTEIN CD8 CHAIN SURFACE ALPHA PRE.	PD02080B 20.69 9.621e-09 538-576
1344	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 9.809e-09 155-178

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1345	BL00282	Kazal serine protease inhibitors family proteins.	BL00282 16.88 6.577e-10 127-149
1345	BL00222	Insulin-like growth factor binding proteins.	BL00222B 11.09 6.940e-10 74-89
1345	BL00621	Tissue factor proteins.	BL00621A 8.69 6.473e-09 5-22
1346	PR00326	GTP1/OBG GTP-BINDING PROTEIN FAMILY SIGNATURE	PR00326A 8.75 1.386e-09 85-105
1346	PF00922	Vesiculovirus phosphoprotein.	PF00922A 19.17 1.724e-09 437-470
1346	PR00449	TRANSFORMING PROTEIN P21 RAS SIGNATURE	PR00449A 13.20 1.931e-09 83-104
1346	PR00905	HYPOTHETICAL MYCOPLASMA LIPOPROTEIN (MG045) SIGNATURE	PR00905H 6.88 5.886e-09 343-363
1348	PR00406	CYTOCHROME B5 REDUCTASE SIGNATURE	PR00406F 3.97 3.520e-10 158-166
1348	PR00014	FIBRONECTIN TYPE III REPEAT SIGNATURE	PR00014B 14.77 2.500e-09 848-858
1348	PD02870	RECEPTOR INTERLEUKIN-1 PRECURSOR.	PD02870B 18.83 3.202e-09 480-512
1348	DM00179	w KINASE ALPHA ADHESION T-CELL.	DM00179 13.97 7.261e-09 205-214
1348	BL00240	Receptor tyrosine kinase class III proteins.	BL00240B 24.70 8.277e-09 263-286
1348	PD02520	RECEPTOR PRECURSOR TRANSMEMBRANE.	PD02520C 10.48 9.203e-09 881-897
1349	PR00698	C.ELEGANS SRG FAMILY INTEGRAL MEMBRANE PROTEIN SIGNATURE	PR00698E 14.43 8.714e-09 97-122
1350	BL00284	Serpins proteins.	BL00284C 28.56 5.714e-32 203-244 BL00284D 16.34 9.640e-19 311-337 BL00284A 15.64 1.783e-18 72-95 BL00284B 17.99 3.045e-16 176-196 BL00284E 19.15 6.250e-14 378-402
1355	PF00023	Ank repeat proteins.	PF00023A 16.03 7.000e-11 69-84 PF00023B 14.20 2.636e-09 131-140
1355	PD00078	REPEAT PROTEIN ANK. NUCLEAR ANKYR.	PD00078B 13.14 2.957e-09 128-140
1355	PF00791	Domain present in ZO-1 and Unc5-like netrin receptors.	PF00791B 28.49 9.587e-09 69-123
1356	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 4.000e-10 339-369
1356	PR00109	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	PR00109D 17.04 4.234e-09 403-425 PR00109B 12.27 1.000e-08 339-357
1358	PR00237	RHODOPSIN-LIKE GPCR SUPERFAMILY SIGNATURE	PR00237G 19.63 3.793e-13 41-67
1358	BL00237	G-protein coupled receptors proteins.	BL00237D 11.23 3.348e-12 51-67
1359	BL00178	Aminoacyl-transfer RNA synthetases class-I proteins.	BL00178B 7.11 3.700e-12 344-354
1360	PF00969	Class II histocompatibility antigen, beta domain proteins.	PF00969A 22.07 5.846e-29 12-54 PF00969B 9.97 6.211e-25 56-91 PF00969C 27.72 7.324e-16 95-144
1361	BL00520	Interleukin-10 family proteins.	BL00520A 6.21 6.471e-09 1-13
1362	BL00520	Interleukin-10 family proteins.	BL00520A 6.21 6.471e-09 1-13
1365	BL00253	Interleukin-1 proteins.	BL00253D 25.67 3.464e-11 95-134
1365	PR00264	INTERLEUKIN-1 SIGNATURE	PR00264C 17.77 3.294e-17 95-123 PR00264B 20.98 6.250e-09 56-82
1366	BL01177	Anaphylatoxin domain proteins.	BL01177E 20.64 4.541e-13 791-817

TABLE 3A

SEQ ID	Database entry ID	Description	Result*
1366	BL00477	Alpha-2-macroglobulin family thiolester region proteins.	BL00477J 19.04 7.207e-29 1221-1251 BL00477F 17.34 8.500e-25 786-815 BL00477G 19.43 8.826e-23 963-994 BL00477A 13.50 9.800e-23 122-150 BL00477L 23.51 8.800e-22 1417-1449 BL00477K 17.42 4.529e-14 1362-1385 BL00477E 17.53 6.538e-13 756-776 BL00477B 9.05 6.625e-13 209-221 BL00477I 18.76 2.650e-12 1065-1091 BL00477D 12.73 4.073e-12 730-739 BL00477H 9.07 5.395e-12 1034-1045 BL00477C 15.70 1.161e-10 236-252
1366	BL00115	Eukaryotic RNA polymerase II heptapeptide repeat proteins.	BL00115V 21.32 5.745e-09 1402-1451
1366	BL00713	Sodium:dicarboxylate symporter family proteins.	BL00713F 16.13 8.989e-09 917-958
1368	BL00983	Ly-6 / u-PAR domain proteins.	BL00983C 12.69 8.714e-16 90-105 BL00983B 8.19 2.161e-10 23-32
1368	BL00272	Snake toxins proteins.	BL00272C 8.27 9.791e-09 94-105

* Results include in order: accession number subtype; raw score; p-value; position of signature in amino acid sequence

359

TABLE 3B

SEQ ID	Database entry ID	Description	Result
685	IPB001400	Somatotropin hormone family	IPB001400A 14.85 1.90e-13 35-58
686	IPB001400	Somatotropin hormone family	IPB001400B 23.62 9.25e-24 79-115 IPB001400A 14.85 4.33e-14 35-58
686	PR00836	Somatotropin hormone family signature I	PR00836A 15.53 1.96e-11 79-92 PR00836B 17.50 9.31e-11 101-119 IPB001400C 13.76 6.28e-10 135-151
688	IPB001400	Somatotropin hormone family	IPB001400B 23.62 1.90e-28 79-115 IPB001400A 14.85 4.91e-16 35-58
688	PR00836	Somatotropin hormone family signature II	PR00836B 17.50 1.43e-15 101-119 PR00836A 15.53 2.35e-13 79-92 IPB001400C 13.76 4.72e-10 135-151
689	IPB000215	Serpins	IPB000215E 15.36 5.76e-17 373-397 IPB000215A 13.01 3.42e-15 77-100 IPB000215D 15.35 8.05e-11 294-320 IPB000215B 9.87 6.04e-10 162-174 IPB000215C 13.90 7.97e-10 189-203
690	PR00390	Phospholipase C signature I	PR00390A 14.24 6.34e-20 191-209
690	IPB002048	EF-hand family	IPB002048 7.91 3.84e-09 43-55
691	IPB000734	Lipase	IPB000734 10.25 8.50e-09 435-449
693	PR00573	Interleukin 8B receptor signature III	PR00573C 9.83 2.15e-09 38-46
693	PR00427	Interleukin-8 receptor signature I	PR00427A 15.48 4.46e-09 34-48
694	IPB000407	GDA1/CD39 family of nucleoside phosphatase	IPB000407C 15.11 4.09e-19 217-239 IPB000407D 11.44 4.27e-15 248-261 IPB000407A 11.93 1.62e-11 101-112 IPB000407B 8.75 2.70e-11 175-186 IPB000407G 17.95 2.80e-11 460-474 IPB000407F 16.53 8.54e-10 430-444
695	PR00237	Rhodopsin-like GPCR superfamily signature VI	PR00237F 14.34 3.20e-09 239-263
695	PR01066	P2Y4 purinoceptor signature II	PR01066B 4.51 6.03e-09 111-126
696	IPB001304	C-type lectin domain	IPB001304A 17.98 3.00e-17 168-192
696	PR01408	Macrophage scavenger receptor signature VI	PR01408F 9.76 4.87e-09 83-107
698	IPB000407	GDA1/CD39 family of nucleoside phosphatase	IPB000407C 15.11 3.30e-16 165-187 IPB000407D 11.44 9.59e-15 196-209 IPB000407B 8.75 9.68e-12 123-134 IPB000407A 11.93 4.50e-10 48-59 IPB000407F 16.53 7.57e-10 377-391
700	IPB000433	ZZ Zinc finger	IPB000433 14.10 4.60e-11 184-200
700	PR00608	Class II cytochrome C signature I	PR00608A 12.75 8.07e-10 118-141
700	IPB000102	Neuraxin / MAP1B repeat	IPB000102A 10.50 5.59e-09 116-144
700	IPB002989	Mycobacterial pentapeptide repeats	IPB002989B 10.80 5.76e-09 110-135
700	PR01286	Orphan nuclear receptor NOR1 signature V	PR01286E 5.27 7.14e-09 133-154
700	PR00456	Ribosomal protein P2 signature V	PR00456E 3.08 8.64e-09 123-137
700	IPB001119	S-layer protein (SLH domain)	IPB001119B 14.79 9.28e-09 115-127 PR00456E 3.08 9.69e-09 122-136
700	IPB001005	Myb DNA binding domain	IPB001005A 11.39 9.71e-09 231-251
701	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 1.00e-09 280-294
701	PR01217	Proline rich extensin signature VIII	PR01217H 5.61 1.67e-09 309-321
702	IPB000345	Cytochrome c family heme-binding site	IPB000345 9.03 7.19e-09 107-119
703	IPB001251	Cellular retinaldehyde-binding protein (CRAL)/Triple function domain (TRIO)	IPB001251A 7.40 5.05e-12 38-49 IPB001251B 14.78 7.14e-12 195-209

360

TABLE 3B

703	PR00180	Cellular retinaldehyde-binding protein signature I	PR00180A 11.19 6.24e-11 37-59 PR00180D 13.13 1.92e-09 202-221
704	IPB002610	Rhomboid family	IPB002610C 5.81 3.81e-10 284-294 IPB002610B 5.33 6.81e-09 225-235
705	PR01256	Otx1 transcription factor signature II	PR01256B 5.92 5.97e-11 221-233 PR01256B 5.92 7.51e-11 218-230 PR01256B 5.92 2.35e-10 219-231 PR01256B 5.92 2.11e-09 220-232 PR01256B 5.92 2.31e-09 222-234 PR01256B 5.92 2.62e-09 217-229
705	IPB001541	SUR2-type hydroxylase/desaturase catalytic domain	IPB001541B 11.65 3.14e-09 223-232 IPB001541B 11.65 3.14e-09 224-233 IPB001541B 11.65 3.14e-09 225-234 IPB001541B 11.65 6.57e-09 222-231
705	PR00910	Luteovirus ORF6 protein signature I	PR00910A 2.74 9.04e-09 756-768
706	IPB001124	Lipid-binding serum glycoprotein	IPB001124D 21.85 2.50e-12 251-287 IPB001124C 25.71 5.08e-11 184-227
707	IPB002495	Glycosyltransferase family 8	IPB002495B 11.16 4.77e-09 273-283
708	IPB001781	LIM domain	IPB001781 11.42 8.77e-11 31-41
710	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442F 15.05 1.00e-40 1624-1667 IPB001442C 14.98 4.82e-40 1537-1571 IPB001442A 26.12 4.09e-39 1298-1350 IPB001442A 26.12 5.40e-35 114-166 IPB001442D 15.34 1.00e-34 1572-1603 IPB001442A 26.12 7.11e-29 799-851 IPB001442A 26.12 1.47e-28 781-833 IPB001442A 26.12 3.48e-28 790-842 IPB001442A 26.12 4.57e-28 814-866
710	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 1.93e-27 1339-1392 IPB000885B 19.15 2.24e-27 783-836 IPB001442A 26.12 2.53e-27 683-735 IPB001442A 26.12 3.59e-27 796-848 IPB000885B 19.15 4.26e-27 780-833 IPB001442A 26.12 4.81e-27 925-977 IPB001442A 26.12 5.
710	IPB001073	Complement C1q protein	IPB001073A 22.14 9.18e-19 1413-1447 IPB000885A 11.46 9.29e-19 744-781 IPB000885B 19.15 9.40e-19 1348-1401 IPB000885B 19.15 9.40e-19 1412-1465 IPB001442A 26.12 9.42e-19 538-590 IPB001442A 26.12 9.42e-19 1304-1356 IPB000885B 19
710	IPB000817	Prion protein	IPB000817A 8.34 7.23e-10 777-819 IPB000885A 11.46 7.26e-10 1064-1101 IPB001442B 12.38 7.30e-10 735-755 IPB001442B 12.38 7.30e-10 938-958 IPB001442B 12.38 7.30e-10 962-982 IPB001442A 26.12 7.36e-10 582-634 IPB001073A 22.14 7.4
710	IPB001285	Synaptophysin/synaptoporin	IPB001285F 6.39 4.08e-09 1379-1423 IPB000885B 19.15 4.11e-09 462-515 IPB000885B 19.15 4.11e-09 1087-1140 IPB001442B 12.38 4.28e-09 103-123 IPB000885A 11.46 4.31e-09 612-649 IPB000885B 19.15 4.35e-09 1213-1266 IPB001442B 12.38
710	IPB003778	DUF183	IPB003778B 27.11 7.31e-09 302-344 IPB001442B 12.38 7.32e-09 794-814

361
TABLE 3B

			IPB001442A 26.12 7.34e-09 629-681 IPB000885B 19.15 7.38e-09 598-651 IPB001442A 26.12 7.42e-09 444-496 IPB001073A 22.14 7.47e-09 975-1009 IPB000885B 19.15 7.5
710	IPB003531	Short hematopoietin receptor family 1	IPB003531C 15.87 9.76e-09 518-535 IPB000817A 8.34 9.81e-09 309-351 IPB000885B 19.15 9.84e-09 1451-1504 IPB001442B 12.38 9.88e-09 302-322 IPB000817A 8.34 9.91e-09 1026-1068 IPB000885B 19.15 1.00e-08 658-711
711	PR00261	Low density lipoprotein (LDL) receptor signature II	PR00261B 15.12 4.13e-22 1101-1122 PR00261C 18.72 2.87e-21 1015-1036 PR00261B 15.12 4.46e-21 1015-1036 PR00261E 18.62 5.74e-21 1144-1165 PR00261B 15.12 1.32e-20 3523-3544
711	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033D 30.18 2.03e-20 2057-2095 PR00261B 15.12 2.61e-20 892-913 PR00261A 15.49 2.73e-20 1053-1074 PR00261D 16.87 6.40e-20 892-913 PR00261B 15.12 6.46e-20 1053-1074 PR00261F 15.46 7.92e-20 892-913 PR00261D 16.87 8.56e-20 3
711	IPB002172	Low density lipoprotein (LDL)- receptor class A (LDLRA) domain	IPB002172 7.37 1.00e-16 2818-2830 PR00261F 15.46 2.10e-16 1185-1206 PR00261D 16.87 2.15e-16 3721-3742 PR00261A 15.49 2.38e-16 2729-2750 PR00261D 16.87 2.38e-16 933-954 PR00261E 18.62 2.97e-16 2729-2750 PR00261F 15.46 3.41e-16
711	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 6.14e-16 206-221 PR00261C 18.72 7.57e-16 2729-2750 PR00261A 15.49 7.92e-16 3562-3583 PR00261F 15.46 8.02e-16 2729-2750 PR00261C 18.72 8.30e-16 3600-3621 IPB000033A 21.82 8.33e-16 2731-2753 PR00261B 15.12 8.53e-16 933-954 PR00261F 15.46 8.68e-16 3562-3583 PR00261C 18.72 9.27e-16 80-101 PR00261F 15.46 9.56e-16 3404-3425 PR00261E 18.62 9.72e-16 3562-3583 PR00261C 18.72 9.76e-16 2938-2959 PR00261E 18.62 1.53e-15 3404-3425 PR00261E 18.62 1.53e-15 3484-3505 PR00261D 16.87 1.63e-15 3641-3662 PR00261C 18.72 1.68e-15 3484-3505 PR00261E 18.62 1.79e-15 3809-3830 PR00261D 16.87 1.84e-15 3364-3385 PR00261A 15.49 2.29e-15 2767-2788 IPB002172 7.37 2.64e-15 89-101 PR00261F 15.46 2.80e-15 2687-2708 PR00261E 18.62 3.12e-15 3523-3544 PR00261C 18.72 3.25e-15 3641-3662
711	PR00764	Complement C9 signature II	PR00764B 12.47 3.36e-15 1048-1068 IPB002172 7.37 3.45e-15 1110-1122 PR00261B 15.12 3.74e-15 3600-3621 PR00261B 15.12 4.33e-15 2893-2914 PR00261C 18.72 4.60e-15 2687-2708

362

TABLE 3B

			IPB000033C 11.58 4.81e-15 3128-3142 IPB000033D 30.18 5.
711	IPB001774	Delta serrate ligand	IPB001774D 19.23 9.89e-14 4240-4286 IPB002172 7.37 1.00e-13 2902-2914 PR00261E 18.62 1.00e-13 2558-2579 PR00261D 16.87 1.53e-13 1185-1206 PR00261C 18.72 1.96e-13 125-146 PR00261F 15.46 2.19e-13 2558-2579 IPB000033C 11.58 2.29e-13 1376-1390 PR00261B 15.12 2.53e-13 2558-2579 IPB002172 7.37 2.59e-13 1062-1074 IPB002172 7.37 2.59e-13 2947-2959 IPB002172 7.37 3.12e-13 2861-2873 PR00764B 12.47 3.38e-13 3636-3656 IPB002172 7.37 3.65e-13 3650-3662
711	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 4.00e-13 206-217 PR00261A 15.49 4.60e-13 1185-1206 IPB000152 8.86 5.09e-13 3019-3034 PR00261B 15.12 5.25e-13 3444-3465 PR00261E 18.62 5.61e-13 125-146 IPB002172 7.37 5.76e-13 2776-2788 IPB002172 7.37 6.29e-13
711	PR00010	Type II EGF-like signature III	PR00010C 6.98 8.13e-11 211-221 IPB002172 7.37 8.43e-11 3532-3544 IPB000033A 21.82 8.71e-11 1187-1209 IPB000033C 11.58 9.00e-11 1774-1788 IPB000152 8.86 9.04e-11 2979-2994 PR00261C 18.72 9.18e-11 2558-2579 IPB000033C 11.58 5.86e-10 2081-2095
711	PR00907	Thrombomodulin signature II	PR00907B 11.50 6.04e-10 4218-4234 IPB000033C 11.58 6.40e-10 411-425 IPB002172 7.37 6.54e-10 942-954 IPB000033C 11.58 6.58e-10 1466-1480 IPB000033A 21.82 7.26e-10 2560-2582 PR00261C 18.72 7.67e-10 2893-2914 PR00010C 6.98 8.55e-10 3024-3034 PR00764B 12.47 8.62e-10 120-140 PR00764B 12.47 8.73e-10 2804-2824 PR00764B 12.47 8.85e-10 3439-3459 IPB002172 7.37 9.31e-10 3493-3505 IPB000033C 11.58 9.46e-10 3084-3098 IPB002172 7.37 1.00e-09 2647-2659 PR00764B 12.47 1.22e-09 3479-3499 IPB000033C 11.58 1.48e-09 736-750 PR00764B 12.47 1.65e-09 2594-2614 IPB000033B 7.05 2.42e-09 3024-3034 PR00764B 12.47 2.63e-09 75-95
711	IPB000970	"Developmental signaling protein, Wnt-1 family"	IPB000970F 23.43 4.19e-09 4241-4289 IPB000033C 11.58 4.21e-09 2404-2418
711	PR00873	Echinoidea (sea urchin) metallothionein signature IV	PR00873D 8.25 4.88e-09 4326-4344 PR00764B 12.47 5.23e-09 2933-2953 IPB000033D 30.18 5.37e-09 4044-4082 PR00764B 12.47 5.66e-09 2553-2573 PR00764B 12.47 5.99e-09 3518-3538 IPB001881B 12.28 6.87e-09 2979-2990
711	IPB001169	"Integrin beta, C-terminus"	IPB001169K 27.45 6.96e-09 2547-2589 IPB000033C 11.58 7.91e-09 1647-1661
711	IPB002557	Chitin binding domain	IPB002557B 12.64 7.92e-09 1236-1249

363

TABLE 3B

			IPB000033C 11.58 8.07e-09 324-338 IPB000033C 11.58 8.07e-09 367-381 IPB000033C 11.58 8.23e-09 1329-1343 IPB001774C 18.25 8.26e-09 4301-4343 PR00010C 6.98 8.46e-09 3900-3910
711	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.62e-09 206-225 PR00764B 12.47 8.70e-09 1096-1116 IPB000033B 7.05 8.82e-09 2984-2994 IPB003886E 12.94 8.88e-09 4100-4110 PR00764B 12.47 9.46e-09 2847-2867
711	IPB000118	Granulin	IPB000118C 7.41 9.65e-09 3822-3863 PR00907B 11.50 9.66e-09 162-178
712	PR00261	Low density lipoprotein (LDL) receptor signature II	PR00261B 15.12 7.43e-18 80-101 PR00261D 16.87 7.25e-17 80-101 PR00261E 18.62 3.53e-16 80-101 PR00261F 15.46 5.39e-16 80-101 PR00261A 15.49 6.08e-16 80-101
712	IPB002172	Low density lipoprotein (LDL)- receptor class A (LDLRA) domain	IPB002172 7.37 2.64e-15 89-101 PR00261C 18.72 3.47e-15 80-101 PR00261A 15.49 7.64e-15 125-146 PR00261F 15.46 8.80e-15 125-146 PR00261D 16.87 1.98e-14 125-146
712	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033A 21.82 3.53e-14 82-104 PR00261C 18.72 1.96e-13 125-146 PR00261E 18.62 5.61e-13 125-146 IPB002172 7.37 6.40e-12 134-146 PR00261B 15.12 9.37e-12 125-146
712	PR00764	Complement C9 signature II	PR00764B 12.47 8.62e-10 120-140
712	PR00907	Thrombomodulin signature II	PR00907B 11.50 9.66e-09 162-178 PR00764B 12.47 1.00e-08 75-95
713	IPB003164	Alpha adaptin carboxyl-terminal domain	IPB003164M 10.25 8.22e-09 164-195
714	PR00205	Cadherin signature VI	PR00205F 19.57 3.86e-16 741-767 PR00205F 19.57 2.13e-15 301-327 PR00205B 20.09 7.30e-15 996-1025 PR00205B 20.09 9.70e-15 250-279 PR00205B 20.09 1.84e-14 475-504 PR00205D 12.22 4.12e-14 332-351
714	IPB002126	Cadherin domain	IPB002126B 12.04 4.79e-14 238-255 PR00205B 20.09 4.94e-14 1210-1239 PR00205B 20.09 7.19e-14 1315-1344 PR00205D 12.22 9.31e-14 1294-1313 IPB002126B 12.04 3.57e-13 463-480 PR00205F 19.57 4.90e-13 1368-1394 IPB002126B 12.04 5.29
715	PR00205	Cadherin signature VI	PR00205F 19.57 3.86e-16 741-767 PR00205F 19.57 2.13e-15 301-327 PR00205B 20.09 7.30e-15 996-1025 PR00205B 20.09 9.70e-15 250-279 PR00205B 20.09 1.84e-14 475-504 PR00205D 12.22 4.12e-14 332-351
715	IPB002126	Cadherin domain	IPB002126B 12.04 4.79e-14 238-255 PR00205B 20.09 4.94e-14 1210-1239 PR00205B 20.09 7.19e-14 1315-1344 PR00205D 12.22 9.31e-14 1294-1313 IPB002126B 12.04 3.57e-13 463-480 PR00205F 19.57 4.90e-13 1368-1394 IPB002126B 12.04 5.29

364

TABLE 3B

716	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469I 10.99 4.86e-16 719-737 IPB002469H 21.17 6.14e-16 674-709 IPB002469J 8.97 3.52e-12 801-817
716	IPB002471	Prolyl endopeptidase family serine active site	IPB002471B 24.90 3.66e-11 706-737 IPB002469G 26.76 9.24e-11 629-667
717	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 1.00e-21 156-181 IPB000822 14.67 4.75e-19 324-349 IPB000822 14.67 4.46e-18 212-237 IPB000822 14.67 3.57e-17 184-209 IPB000822 14.67 7.43e-17 240-265 IPB000822 14.67 1.00e-16 296-321 IPB000822 14.67 2.69e-15 62-87 IPB000822 14.67 4.38e-15 352-377
717	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 8.20e-15 181-194
717	IPB001275	DM DNA binding domain	IPB001275 19.17 9.07e-15 172-211 PR00048A 9.94 3.77e-14 321-334 PR00048A 9.94 8.62e-14 349-362 PR00048A 9.94 3.57e-13 153-166 IPB001275 19.17 9.71e-13 144-183 IPB000822 14.67 1.95e-12 268-293 PR00048A 9.94 2.06e-12 237-250 PR00048A 9.94 4.18e-12 209-222 IPB000822 14.67 9.53e-12 34-59 PR00048A 9.94 6.21e-11 265-278 IPB001275 19.17 8.71e-11 200-239 PR00048A 9.94 1.41e-10 293-306 IPB001275 19.17 4.16e-10 312-351 PR00048B 5.52 5.50e-10 197-206 PR00048A 9.94 7.55e-10 59-72 PR00048B 5.52 9.36e-10 337-346 PR00048B 5.52 1.00e-09 169-178 PR00048B 5.52 3.50e-09 225-234 IPB001275 19.17 3.62e-09 256-295 PR00048B 5.52 4.50e-09 365-374 IPB001275 19.17 5.22e-09 228-267 IPB001275 19.17 8.75e-09 284-323
718	IPB000221	Protamine P1	IPB000221 5.48 2.97e-12 74-100 IPB000221 5.48 9.30e-12 63-89 IPB000221 5.48 2.19e-11 103-129 IPB000221 5.48 2.59e-11 64-90 IPB000221 5.48 3.91e-11 78-104
718	IPB000492	Protamine 2 (PRM2)	IPB000492B 5.26 5.88e-11 98-132 IPB000221 5.48 6.16e-11 92-118 IPB000221 5.48 6.43e-11 99-125 IPB000221 5.48 7.62e-11 60-86 IPB000492B 5.26 9.35e-11 79-113 IPB000492B 5.26 9.35e-11 102-136 IPB000221 5.48 2.73e-10 118-144 IPB000221 5.48 4.70e-10 62-88 IPB000221 5.48 4.70e-10 94-120 IPB000492B 5.26 6.97e-10 103-137 IPB000492B 5.26 8.12e-10 106-140 IPB000492B 5.26 8.53e-10 105-139 IPB000221 5.48 8.89e-10 101-127 IPB000492B 5.26 9.06e-10 78-112 IPB000492B 5.26 9.69e-10 100-134 IPB000221 5.48 1.00e-09 83-109 IPB000221 5.48 1.46e-09 65-91

365
TABLE 3B

			IPB000221 5.48 3.31e-09 109-135 IPB000221 5.48 3.31e-09 122-148 IPB000492B 5.26 3.84e-09 75-109 IPB000221 5.48 5.15e-09 107-133 IPB000221 5.48 5.27e-09 52-78
718	PR00055	HIV TAT domain signature III	PR00055C 9.12 5.92e-09 16-32 IPB000221 5.48 6.19e-09 116-142 IPB000492B 5.26 6.38e-09 94-128 IPB000492B 5.26 6.67e-09 107-141 IPB000221 5.48 6.88e-09 97-123 IPB000221 5.48 6.88e-09 111-137 IPB000492B 5.26 7.75e-09 77-111 IPB000492B 5.26 8.34e-09 65-99
718	IPB000271	Ribosomal protein L34	IPB000271 15.87 9.78e-09 111-148 IPB000221 5.48 9.88e-09 124-150 IPB000492B 5.26 9.90e-09 111-145 IPB000221 5.48 1.00e-08 76-102
720	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 6.54e-17 2348-2363 IPB000152 8.86 4.18e-15 2191-2206 IPB000152 8.86 3.84e-14 2232-2247 IPB000152 8.86 3.86e-13 2108-2123
720	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 4.78e-13 2232-2251
720	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 5.50e-13 2191-2202
720	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.29e-13 1028-1065
720	PR00010	Type II EGF-like signature III	PR00010C 6.98 9.47e-13 2353-2363 IPB003006B 20.23 1.00e-12 1119-1156
720	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 3.70e-12 2196-2206 IPB001881B 12.28 5.20e-12 2348-2359
720	IPB002861	Reeler domain	IPB002861B 10.50 6.52e-12 1435-1463 IPB002861B 10.50 7.12e-12 1606-1634
720	PR01303	Plasmodium circumsporozoite protein signature IV	PR01303D 10.57 7.20e-12 1441-1458 PR00010C 6.98 1.75e-11 2196-2206 IPB000152 8.86 1.96e-11 2023-2038 IPB001881B 12.28 4.79e-11 2232-2243 IPB003006B 20.23 4.91e-11 386-423 IPB003006B 20.23 5.30e-11 1208-1245 IPB002861B 10.50 7.08e-11 1549-1577 IPB003006B 20.23 8.43e-11 199-236 IPB001881B 12.28 8.58e-11 2066-2077 IPB001881B 12.28 9.53e-11 2023-2034 IPB003006B 20.23 9.61e-11 756-793
720	IPB000981	Neurohypophysial hormone	IPB000981A 17.34 1.60e-10 1594-1621 IPB003006B 20.23 2.08e-10 847-884 IPB003886D 13.91 2.33e-10 2191-2210 IPB000033B 7.05 4.48e-10 2353-2363
720	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 5.83e-10 2116-2136 PR01303D 10.57 5.90e-10 1612-1629 IPB000033B 7.05 7.10e-10 2113-2123
720	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862A 12.54 8.02e-10 1714-1729
720	PR00907	Thrombomodulin signature VII	PR00907G 10.43 8.09e-10 2348-2374 IPB003006B 20.23 8.56e-10 104-141 IPB001881B 12.28 8.71e-10 2108-2119 PR00907G 10.43 8.85e-10 2232-2258 IPB003006B 20.23 8.92e-10 938-975 IPB003886D 13.91 9.41e-10 2348-2367

366

TABLE 3B

			PR00907B 11.50 9.64e-10 2228-2244 IPB003006B 20.23 1.35e-09 479-516 PR01303D 10.57 2.00e-09 1726-1743
720	PR01472	Intercellular adhesion molecule/vascular cell adhesion molecule-1 signature III	PR01472C 14.40 3.41e-09 994-1009 IPB003886D 13.91 3.49e-09 2108-2127
720	IPB000561	EGF-like domain	IPB000561 4.89 3.57e-09 2357-2365 PR00010C 6.98 3.63e-09 2113-2123 IPB003006B 20.23 3.77e-09 1299-1336 IPB000033A 21.82 4.35e-09 2053-2075 IPB002861B 10.50 4.48e-09 1663-1691 IPB003006B 20.23 4.81e-09 10-47 IPB003367A 11.78 5.13e-09 2318-2338 IPB003006B 20.23 5.50e-09 572-609
720	PR01474	Vascular cell adhesion molecule-1 (VCAM-1) signature VI	PR01474F 14.81 5.76e-09 1221-1234 IPB003006B 20.23 5.85e-09 293-330
720	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 5.85e-09 393-416 PR01536C 19.92 6.08e-09 1126-1149 PR01536C 19.92 7.46e-09 763-786 PR01536C 19.92 7.58e-09 1215-1238 PR00010C 6.98 8.02e-09 2237-2247 IPB001862A 12.54 8.55e-09 1486-1501 IPB002861B 10.50 8.98e-09 1720-1748 IPB002861C 23.17 9.02e-09 1650-1704
720	IPB000967	Zinc finger NF-X1 type	IPB000967E 21.88 9.20e-09 1443-1483
720	IPB000118	Granulin	IPB000118B 7.94 9.20e-09 2011-2049 PR00907G 10.43 9.27e-09 2108-2134 PR00907B 11.50 9.43e-09 2344-2360 IPB002861B 10.50 9.59e-09 1492-1520
721	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 8.05e-14 71-95 IPB000135D 2.13 5.27e-13 72-96 IPB000135D 2.13 9.46e-12 73-97 IPB000135D 2.13 4.78e-11 70-94
721	IPB003874	CDC45-like protein	IPB003874C 5.49 8.27e-11 74-85
721	IPB000897	GTP-binding signal recognition particle (SRP54) domain	IPB000897A 9.15 8.60e-11 454-473 IPB000135D 2.13 3.05e-10 74-98
721	IPB001580	Calreticulin family	IPB001580F 2.93 8.31e-10 78-87 IPB000135D 2.13 9.02e-10 69-93 IPB000135D 2.13 1.00e-09 65-89 IPB001580F 2.93 1.45e-09 76-85 IPB000135D 2.13 5.09e-09 66-90 IPB001580F 2.93 6.85e-09 74-83 IPB000135D 2.13 7.00e-09 75-99 IPB000135D 2.13 8.00e-09 68-92 IPB000135D 2.13 9.36e-09 63-87
722	IPB001140	ABC transporter transmembrane region	IPB001140A 21.73 8.36e-20 1311-1357 IPB001140A 21.73 9.29e-18 499-545 IPB001140B 15.62 4.79e-15 615-653 IPB001140B 15.62 1.16e-10 1427-1465
722	PR00326	GTP1/OBG GTP-binding protein family signature I	PR00326A 8.70 6.66e-10 513-533
722	IPB000795	GTP-binding elongation factor	IPB000795A 10.67 7.88e-10 1324-1339
722	IPB000897	GTP-binding signal recognition particle (SRP54) domain	IPB000897A 9.15 1.54e-09 512-531 IPB000795A 10.67 2.85e-09 512-527 PR00326A 8.70 4.49e-09 1325-1345 IPB000897A 9.15 5.57e-09 1324-1343
722	IPB001324	Phosphoribulokinase family	IPB001324A 18.12 8.00e-09 1321-1342

367

TABLE 3B

722	PR00364	Disease resistance protein signature I	PR00364A 8.29 8.00e-09 512-527
722	PR01014	Neuropeptide Y2 receptor signature VI	PR01014F 15.22 8.74e-09 647-663
723	PR01217	Proline rich extensin signature VII	PR01217G 4.02 7.16e-09 242-267 PR01217D 4.57 7.49e-09 495-516
723	IPB001084	Microtubule associated Tau protein	IPB001084C 7.66 9.64e-09 308-325
723	IPB001101	Plectin repeat	IPB001101K 8.53 9.92e-09 29-72
724	IPB001552	Acyl-CoA dehydrogenase	IPB001552E 22.77 2.46e-19 158-198 IPB001552D 24.88 5.35e-19 67-109 IPB001552C 25.04 7.75e-15 13-53
725	IPB000998	MAM domain	IPB000998D 18.66 1.96e-15 526-549
725	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.77e-15 236-255
725	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 2.89e-14 109-124
725	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 5.00e-14 191-202 IPB000152 8.86 1.00e-13 236-251 IPB000152 8.86 1.82e-13 191-206 IPB001881B 12.28 4.75e-13 109-120
725	IPB001774	Delta serrate ligand	IPB001774C 18.25 9.13e-13 71-113 IPB000998B 17.20 1.00e-12 409-421
725	PR00020	MAM domain signature I	PR00020A 20.48 2.88e-11 407-425 IPB000998C 18.63 5.30e-11 463-478 IPB001881B 12.28 8.58e-11 236-247
725	PR00907	Thrombomodulin signature II	PR00907B 11.50 2.44e-10 143-159
725	IPB000561	EGF-like domain	IPB000561 4.89 3.25e-10 80-88
725	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 5.35e-10 241-251 IPB000033B 7.05 5.97e-09 196-206
725	IPB000167	Dehydrin	IPB000167A 8.58 7.14e-09 323-350
725	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 9.79e-09 158-178
726	IPB001258	NHL repeat	IPB001258B 28.61 4.30e-17 619-653 IPB001258B 28.61 7.00e-17 525-559 IPB001258B 28.61 1.27e-16 431-465 IPB001258B 28.61 5.91e-16 478-512
726	PR01406	B-box zinc finger signature I	PR01406A 20.90 8.36e-12 112-129 IPB001258B 28.61 5.60e-11 572-606
726	IPB003649	B-Box C-terminal domain	IPB003649B 22.16 3.68e-10 115-134
726	IPB001869	Thiol-activated cytolytins	IPB001869C 15.61 6.06e-09 396-419
727	IPB000198	RhoGAP domain	IPB000198C 16.49 8.31e-16 923-940 IPB000198B 12.47 9.10e-15 833-850
727	IPB002219	Phorbol esters/diacylglycerol binding domain	IPB002219B 12.53 3.89e-11 724-739 IPB000198A 15.95 9.61e-10 781-797
727	IPB002551	Coronavirus S1 glycoprotein	IPB002551J 18.56 3.60e-09 470-511
727	IPB001369	Purine and other phosphorylases family 2	IPB001369C 24.81 4.27e-09 36-76
727	IPB003351	Dishevelled specific domain	IPB003351C 13.82 7.24e-09 1025-1064
729	IPB002870	Reprolysin family propeptide	IPB002870B 24.73 6.23e-24 131-169 IPB002870F 18.81 6.54e-16 456-480
729	IPB001762	Disintegrin	IPB001762A 23.93 6.50e-15 359-399 IPB002870E 11.90 8.67e-14 414-426 IPB002870D 16.31 8.77e-13 383-398
729	PR01303	Plasmodium circumsporozoite protein signature IV	PR01303D 10.57 1.42e-11 1173-1190 PR01303D 10.57 1.40e-10 1488-1505 IPB002870A 12.22 2.29e-10 81-97 IPB002870C 11.01 2.80e-10 344-354 PR01303D 10.57 3.91e-10 1098-1115
729	IPB000130	"Neutral zinc metallopeptidases, zinc-binding region"	IPB000130 5.86 7.19e-10 412-422

368

TABLE 3B

729	IPB000118	Granulin	IPB000118G 12.18 4.31e-09 1471-1519
729	IPB002861	Reeler domain	IPB002861C 23.17 5.34e-09 969-1023 PR01303D 10.57 6.50e-09 1229-1246 IPB002861B 10.50 7.75e-09 1223-1251
729	PR00269	Pleiotrophin/midkine family signature I	PR00269A 12.42 9.33e-09 1162-1186
730	PR01478	Leukotriene B4 type 2 receptor signature V	PR01478E 5.85 7.56e-10 149-177
735	IPB001331	Guanine-nucleotide dissociation stimulators CDC24 family	IPB001331C 16.09 7.35e-14 302-327
737	IPB002004	"Poly-adenylate binding protein, unique domain"	IPB002004C 13.84 8.14e-10 189-231
741	PR01276	Type II keratin signature II	PR01276B 9.79 9.27e-10 147-159
742	PR00205	Cadherin signature II	PR00205B 20.09 5.95e-20 252-281 PR00205D 12.22 3.25e-16 654-673 PR00205B 20.09 7.60e-15 142-171 PR00205F 19.57 1.00e-14 520-546 PR00205G 13.05 1.37e-13 657-674 PR00205F 19.57 3.10e-13 623-649 PR00205D 12.22 5.80e-13 231-250 PR00205D 12.22 5.80e-13 551-570 PR00205B 20.09 6.40e-13 469-498
742	IPB002126	Cadherin domain	IPB002126B 12.04 8.71e-13 560-577 PR00205F 19.57 1.26e-12 308-334 PR00205G 13.05 1.30e-12 340-357 PR00205G 13.05 4.90e-12 554-571 PR00205D 12.22 5.37e-12 337-356 PR00205D 12.22 8.20e-12 448-467 PR00205G 13.05 8.50e-12 234-251 PR00205G 13.05 6.84e-11 451-468 IPB002126B 12.04 7.43e-11 240-257 PR00205F 19.57 7.63e-11 417-443 PR00205A 17.38 8.56e-11 301-320 IPB002126B 12.04 3.03e-10 457-474 IPB002126B 12.04 9.42e-10 130-147 IPB002126A 14.68 3.67e-09 312-328 PR00205A 17.38 4.71e-09 513-532 PR00205E 10.82 5.50e-09 570-583 IPB002126A 14.68 6.33e-09 204-220 PR00205C 13.59 6.62e-09 640-652 PR00205B 20.09 7.06e-09 572-601 PR00205D 12.22 8.27e-09 121-140 PR00205G 13.05 9.82e-09 124-141
744	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862C 26.48 8.94e-09 119-167
745	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 6.00e-24 216-241 IPB000822 14.67 9.18e-21 160-185 IPB000822 14.67 1.75e-20 328-353 IPB000822 14.67 4.00e-20 518-543 IPB000822 14.67 8.50e-20 244-269 IPB000822 14.67 9.25e-19 490-515 IPB000822 14.67 7.92e-18 188-213 IPB000822 14.67 9.31e-18 356-381 IPB000822 14.67 9.36e-17 272-297 IPB000822 14.67 3.40e-16 384-409 IPB000822 14.67 8.80e-16 300-325
745	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 5.50e-15 381-394 PR00048A 9.94 1.00e-14 269-282

369

TABLE 3B

			PR00048A 9.94 1.00e-14 543-556 PR00048A 9.94 3.08e-14 185-198 PR00048A 9.94 4.46e-14 487-500 IPB000822 14.67 6.06e-14 440-465 IPB000822 14.67 2.50e-13 412-437 PR00048A 9.94 3.57e-13 297-310 PR00048A 9.94 6.79e-13 213-226 PR00048A 9.94 7.43e-13 409-422 IPB000822 14.67 8.00e-13 132-157
745	IPB001275	DM DNA binding domain	IPB001275 19.17 8.00e-13 148-187 PR00048A 9.94 3.12e-12 241-254 PR00048A 9.94 5.76e-12 515-528 PR00048B 5.52 7.00e-12 173-182 IPB001275 19.17 7.58e-12 204-243 PR00048A 9.94 8.41e-12 353-366 IPB001275 19.17 3.96e-11 506-545 IPB000822 14.67 4.43e-11 546-571 IPB001275 19.17 5.76e-11 176-215 PR00048A 9.94 6.21e-11 325-338 PR00048B 5.52 7.00e-11 341-350 PR00048B 5.52 9.25e-11 503-512 PR00048B 5.52 1.00e-10 229-238 IPB001275 19.17 1.49e-10 344-383 IPB001275 19.17 4.41e-10 316-355
745	IPB001222	TFIIS zinc ribbon domain	IPB001222 24.63 5.16e-10 490-526 IPB001275 19.17 5.50e-10 232-271 PR00048A 9.94 7.14e-10 129-142 PR00048A 9.94 7.14e-10 157-170 PR00048A 9.94 1.38e-09 437-450 IPB001275 19.17 1.46e-09 372-411 IPB001275 19.17 3.39e-09 288-327 PR00048B 5.52 5.50e-09 531-540 IPB001222 24.63 8.35e-09 160-196 IPB001275 19.17 9.09e-09 260-299
745	IPB001142	Yeast membrane protein DUP	IPB001142B 22.92 9.60e-09 290-335
745	IPB002867	Cysteine-rich domain (C6HC)	IPB002867C 19.46 9.76e-09 129-146 PR00048B 5.52 1.00e-08 313-322
746	IPB001909	KRAB box	IPB001909 17.37 8.65e-30 37-71
746	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 6.00e-24 291-316 IPB000822 14.67 9.18e-21 235-260 IPB000822 14.67 1.75e-20 403-428 IPB000822 14.67 8.50e-20 319-344 IPB000822 14.67 7.92e-18 263-288 IPB000822 14.67 9.31e-18 431-456 IPB000822 14.67 9.36e-17 347-372 IPB000822 14.67 3.40e-16 459-484 IPB000822 14.67 8.80e-16 375-400
746	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 5.50e-15 456-469 PR00048A 9.94 1.00e-14 344-357 PR00048A 9.94 3.08e-14 260-273 IPB000822 14.67 6.06e-14 515-540 IPB000822 14.67 2.50e-13 487-512 PR00048A 9.94 3.57e-13 372-385 PR00048A 9.94 6.79e-13 288-301 PR00048A 9.94 7.43e-13 484-497 IPB000822 14.67 8.00e-13 207-232
746	IPB001275	DM DNA binding domain	IPB001275 19.17 8.00e-13 223-262 PR00048A 9.94 3.12e-12 316-329

370

TABLE 3B

			PR00048B 5.52 7.00e-12 248-257 IPB001275 19.17 7.58e-12 279-318 PR00048A 9.94 8.41e-12 428-441 IPB001275 19.17 5.76e-11 251-290 PR00048A 9.94 6.21e-11 400-413 PR00048B 5.52 7.00e-11 416-425 PR00048B 5.52 1.00e-10 304-313 IPB001275 19.17 1.49e-10 419-458 IPB001275 19.17 4.41e-10 391-430 IPB001275 19.17 5.50e-10 307-346 PR00048A 9.94 7.14e-10 204-217 PR00048A 9.94 7.14e-10 232-245 PR00048A 9.94 1.38e-09 512-525 IPB001275 19.17 1.46e-09 447-486 IPB001275 19.17 3.39e-09 363-402
746	IPB001222	TFIIS zinc ribbon domain	IPB001222 24.63 8.35e-09 235-271 IPB001275 19.17 9.09e-09 335-374
746	IPB001142	Yeast membrane protein DUP	IPB001142B 22.92 9.60e-09 365-410
746	IPB002867	Cysteine-rich domain (C6HC)	IPB002867C 19.46 9.76e-09 204-221 PR00048B 5.52 1.00e-08 388-397
747	IPB000348	emp24/gp25L/p24 family	IPB000348B 26.69 5.33e-31 143-188 IPB000348A 15.21 3.63e-12 78-96
748	IPB000560	Histidine acid phosphatase	IPB000560 17.02 1.00e-16 31-53
749	PR00405	HIV Rev interacting protein signature II	PR00405B 10.10 8.29e-19 558-575 PR00405C 18.05 9.55e-19 579-600 PR00405A 18.83 4.00e-18 539-558
749	IPB000906	ZU5 domain	IPB000906G 25.85 4.32e-12 827-875 IPB000906D 23.89 7.43e-09 846-900
751	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.00e-24 753-778
751	IPB001909	KRAB box	IPB001909 17.37 2.86e-21 344-378 IPB000822 14.67 3.57e-17 695-720 IPB000822 14.67 3.25e-14 605-630 IPB000822 14.67 9.44e-14 781-806 IPB000822 14.67 2.50e-13 723-748
751	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 3.37e-11 602-615 PR00048A 9.94 4.32e-11 778-791 PR00048A 9.94 5.26e-11 692-705 IPB000822 14.67 6.14e-11 633-658 PR00048A 9.94 9.53e-11 750-763 PR00048B 5.52 1.00e-10 766-775 PR00048A 9.94 3.86e-10 720-733
751	IPB001580	Calreticulin family	IPB001580F 2.93 1.00e-09 514-523 PR00048A 9.94 6.25e-09 630-643 PR00048B 5.52 6.50e-09 708-717
751	PR01073	Presenilin 1 signature III	PR01073C 1.45 6.62e-09 509-520
751	IPB001275	DM DNA binding domain	IPB001275 19.17 8.18e-09 769-808
751	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 8.45e-09 507-531
753	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 8.11e-14 261-275
753	PR00364	Disease resistance protein signature IV	PR00364D 14.89 4.60e-09 103-119
753	PR00019	Leucine-rich repeat signature II	PR00019B 11.42 8.91e-09 154-167
754	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 7.84e-26 1244-1271 IPB001599F 18.95 7.00e-24 785-814 IPB001599H 18.42 6.40e-20 1019-1046 IPB001599A 10.97 9.69e-18 123-141

371

TABLE 3B

			IPB001599N 24.85 2.24e-14 1437-1469
754	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 1257-1271 IPB001599M 13.29 4.71e-13 1384-1395 IPB001599G 13.87 8.94e-13 987-996 IPB001599B 7.45 4.89e-12 209-221 IPB001599D 11.61 6.90e-12 728-738 IPB001599J 20.99 3.00e-11 1085-1110 IPB001599I 10.83 7.60e-11 1054-1063 IPB001599K 8.15 1.46e-10 1214-1225 IPB001599C 14.40 3.55e-09 236-252 IPB001599E 11.06 9.77e-09 755-764
755	IPB002181	Fibrinogen beta and gamma chains C-terminal globular domain	IPB002181E 27.75 4.44e-21 344-376 IPB002181D 29.18 5.14e-19 298-338 IPB002181F 18.85 2.13e-14 398-421 IPB002181C 15.87 5.78e-12 280-292 IPB002181A 18.44 2.32e-10 244-260
756	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 5.30e-11 457-494
756	PR00014	Fibronectin type III repeat signature IV	PR00014D 15.12 5.26e-10 671-685 IPB003006B 20.23 5.68e-10 174-211 IPB003006B 20.23 5.68e-10 275-312
756	PR00406	Cytochrome B5 reductase signature VI	PR00406F 4.29 6.03e-09 140-148
756	IPB003866	Isoflavone reductase	IPB003866D 19.80 9.48e-09 454-506
757	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 6.85e-13 240-254
757	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 7.14e-11 149-162 PR00019B 11.42 8.00e-10 98-111 PR00019B 11.42 7.55e-09 122-135 PR00019B 11.42 8.09e-09 146-159
757	IPB002889	WSC domain	IPB002889B 11.76 8.97e-09 599-645
757	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 9.31e-09 335-372 IPB002889B 11.76 9.44e-09 598-644
758	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 6.85e-13 240-254
758	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 7.14e-11 149-162 PR00019B 11.42 8.00e-10 98-111 PR00019B 11.42 7.55e-09 122-135 PR00019B 11.42 8.09e-09 146-159
758	IPB002889	WSC domain	IPB002889B 11.76 8.97e-09 603-649
758	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 9.31e-09 335-372 IPB002889B 11.76 9.44e-09 602-648
759	IPB000203	GPS domain	IPB000203A 18.40 9.25e-20 966-996 IPB000203B 13.98 8.88e-15 1086-1107
759	IPB000832	G-protein coupled receptors family 2 (secretin-like)	IPB000832C 19.53 9.46e-13 1086-1115
759	PR00249	Secretin-like GPCR superfamily signature III	PR00249C 15.44 1.73e-10 1088-1111 IPB000832G 15.17 7.81e-09 1256-1281
760	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 4.00e-24 277-302 IPB000822 14.67 3.45e-21 361-386 IPB000822 14.67 1.75e-20 193-218 IPB000822 14.67 3.25e-19 109-134 IPB000822 14.67 4.00e-19 389-414 IPB000822 14.67 8.50e-19 165-190 IPB000822 14.67 1.00e-18 249-274 IPB000822 14.67 5.85e-18 305-330 IPB000822 14.67 1.60e-16 137-162

372
TABLE 3B

			IPB000822 14.67 3.40e-16 333-358 IPB000822 14.67 5.50e-15 221-246
760	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 6.54e-14 330-343
760	IPB001275	DM DNA binding domain	IPB001275 19.17 6.55e-14 237-276 IPB001275 19.17 8.05e-14 321-360 IPB001275 19.17 8.20e-14 153-192 IPB001275 19.17 2.14e-13 349-388 IPB001275 19.17 4.57e-13 265-304 PR00048A 9.94 4.86e-13 218-231 PR00048A 9.94 4.86e-13 274-28
760	IPB002867	Cysteine-rich domain (C6HC)	IPB002867C 19.46 8.11e-09 274-291 PR00048A 9.94 8.12e-09 358-371
760	IPB002634	BolA-like protein	IPB002634A 23.30 8.25e-09 298-332
760	PR00995	36kDa capillovirus serine protease (S35) signature VI	PR00995F 16.50 9.73e-09 311-329
761	PR00121	Sodium/potassium-transporting ATPase signature IV	PR00121D 16.73 7.12e-15 173-194
761	IPB001757	E1-E2 ATPases	IPB001757B 13.64 9.65e-13 588-617 IPB001757A 14.16 4.18e-12 179-190
761	PR00119	P-type cation-transporting ATPase superfamily signature II	PR00119B 12.03 9.61e-12 180-194
761	IPB000150	Cof protein	IPB000150C 20.72 7.47e-09 595-627
763	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 8.88e-09 172-197
764	IPB001310	HIT (Histidine triad) family	IPB001310A 18.76 3.25e-18 177-207 IPB001310B 21.00 2.93e-12 241-267
764	PR00332	Histidine triad family signature II	PR00332B 14.02 6.26e-10 189-207
767	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 4.52e-10 101-125 IPB000135D 2.13 9.71e-10 103-127 IPB000135D 2.13 9.90e-10 100-124 IPB000135D 2.13 3.18e-09 104-128 IPB000135D 2.13 9.55e-09 102-126
768	PR00074	Protein-lysine 6-oxidase precursor signature V	PR00074E 11.34 9.46e-14 327-347 PR00074B 7.56 4.98e-12 260-284
768	IPB001695	Lysyl oxidase	IPB001695E 9.12 5.70e-12 244-285
768	PR00258	Speract receptor signature IV	PR00258D 14.29 7.39e-12 94-108 PR00258E 14.06 3.38e-11 117-129 PR00258A 13.56 1.54e-10 29-45 PR00074D 21.66 2.94e-10 305-326 PR00258A 13.56 3.70e-10 139-155 PR00258C 9.05 4.95e-10 177-187 PR00258D 14.29 6.29e-10 210-224 PR00258C 9.05 9.34e-10 63-73 PR00258B 7.94 6.14e-09 48-59 IPB001695F 11.10 6.87e-09 285-313
771	IPB001084	Microtubule associated Tau protein	IPB001084C 7.66 1.00e-08 105-122
773	IPB000374	Phosphatidate cytidyltransferase	IPB000374B 15.86 2.06e-27 358-385 IPB000374A 12.59 3.65e-16 254-266
774	PR00320	G protein beta WD-40 repeat signature I	PR00320A 13.15 7.95e-11 190-204 PR00320B 12.82 2.08e-10 190-204 PR00320C 12.32 4.33e-09 190-204
775	IPB001422	Neuromodulin (GAP-43)	IPB001422C 16.82 1.95e-10 155-190
775	IPB001990	Granins (chromogranin or secretogranin)	IPB001990C 33.59 8.01e-10 150-197
776	IPB002549	Domain of unknown function DUF20	IPB002549B 19.59 9.27e-09 229-266
778	IPB002884	Proprotein convertase P-domain	IPB002884B 15.69 6.33e-09 114-131
779	IPB000361	Hypothetical hesB/yadR/yfhF family	IPB000361B 19.14 3.08e-19 119-150 IPB000361A 17.83 2.71e-16 70-90

373

TABLE 3B

780	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 9.28e-10 131-168
783	IPB002223	Pancreatic trypsin inhibitor (Kunitz) family	IPB002223 17.66 3.88e-25 556-590
783	IPB000885	Fibrillar collagen C-terminal domain	IPB000885A 11.46 5.57e-19 13-50
783	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 6.26e-19 6-58 IPB001442A 26.12 4.44e-18 3-55 IPB001442A 26.12 3.17e-17 185-237 IPB001442A 26.12 3.60e-17 191-243 IPB000885B 19.15 5.72e-17 2-55 IPB000885B 19.15 6.29e-17 11-64 IPB001442A 26.12 7.51e-17 12-64 IPB001442A 26.12 1.21e-16 197-249 IPB000885B 19.15 2.19e-16 193-246 IPB001442A 26.12 3.51e-16 9-61 IPB000885A 11.46 5.06e-16 198-235 IPB001442A 26.12 6.02e-16 188-240 IPB000885B 19.15 7.83e-16 8-61 IPB000885A 11.46 1.61e-15 19-56 IPB000885B 19.15 3.65e-15 202-255 IPB000885B 19.15 4.39e-15 184-237 IPB000885B 19.15 4.49e-15 190-243 IPB000885B 19.15 8.09e-15 17-70 IPB001442A 26.12 9.29e-15 182-234 IPB001442A 26.12 9.80e-15 15-67
783	PR00453	Von Willebrand factor type A domain signature I	PR00453A 11.78 1.75e-14 265-282 IPB000885A 11.46 2.29e-14 201-238 IPB000885A 11.46 3.92e-14 210-247 IPB000885B 19.15 6.76e-14 14-67 IPB000885B 19.15 6.97e-14 187-240 IPB000885A 11.46 7.08e-14 22-59 IPB001442A 26.12 7.65e-14 200-252 IPB000885B 19.15 7.78e-14 5-58 IPB001442A 26.12 8.63e-14 203-255 IPB000885A 11.46 9.77e-14 25-62 IPB001442A 26.12 1.00e-13 194-246 IPB000885A 11.46 1.44e-13 10-47 IPB000885A 11.46 2.89e-13 195-232 IPB001442B 12.38 4.67e-13 60-80 IPB000885A 11.46 6.33e-13 207-244 IPB000885B 19.15 7.07e-13 196-249 IPB000885A 11.46 7.33e-13 16-53 IPB000885B 19.15 7.46e-13 199-252 IPB001442B 12.38 1.31e-12 22-42
783	IPB001073	Complement C1q protein	IPB001073A 22.14 1.36e-12 56-90 IPB001073A 22.14 1.72e-12 203-237 IPB001073A 22.14 2.80e-12 119-153 IPB000885A 11.46 2.93e-12 7-44 IPB001442A 26.12 5.05e-12 24-76 IPB000885A 11.46 5.93e-12 213-250 IPB000885A 11.46 6.04e-12 20
783	PR00759	Basic protease (Kunitz-type) inhibitor family signature III	PR00759C 12.43 6.28e-11 575-590 IPB001073A 22.14 7.00e-11 59-93 IPB000885A 11.46 7.57e-11 28-65 IPB001073A 22.14 8.17e-11 142-176 IPB001073A 22.14 8.33e-11 50-84 IPB001073A 22.14 8.67e-11 15-49 IPB001442B 12.38 8.71e-11 37-57

TABLE 3B

783	IPB000817	Prion protein	IPB000817A 8.34 9.70e-10 132-174 PR00759B 12.35 9.72e-10 565-575 IPB001442A 26.12 9.92e-10 30-82 IPB000885A 11.46 1.83e-09 189-226 IPB001442B 12.38 1.97e-09 210-230 IPB001073A 22.14 2.27e-09 128-162 IPB000885B 19.15 2.47e-09
784	IPB001541	SUR2-type hydroxylase/desaturase catalytic domain	IPB001541A 12.30 5.50e-11 164-176 IPB001541B 11.65 4.86e-09 251-260
784	IPB001369	Purine and other phosphorylases family 2	IPB001369A 12.23 8.71e-09 2-15
785	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.96e-10 367-404 IPB003006B 20.23 6.19e-09 1589-1626
786	PR00918	Calicivirus non-structural polyprotein family signature I	PR00918A 13.81 3.59e-12 27-47
786	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 4.25e-12 186-210 IPB000135D 2.13 9.24e-12 187-211 IPB000135D 2.13 6.42e-11 188-212 IPB000135D 2.13 1.68e-10 185-209
786	IPB002078	Sigma-54 factor interaction protein family	IPB002078A 20.43 6.31e-10 33-67
786	PR00364	Disease resistance protein signature I	PR00364A 8.29 7.11e-10 32-47
786	IPB000765	GTP1/OBG family	IPB000765 26.91 7.67e-10 31-74
786	IPB000897	GTP-binding signal recognition particle (SRP54) domain	IPB000897A 9.15 8.26e-10 393-412
786	IPB001580	Calreticulin family	IPB001580F 2.93 8.31e-10 200-209 IPB001580F 2.93 9.44e-10 201-210
786	IPB000623	Shikimate kinase	IPB000623A 19.06 1.64e-09 394-423
786	IPB000619	Guanylate kinase	IPB000619A 18.08 1.86e-09 394-411 IPB001580F 2.93 1.90e-09 199-208
786	PR00094	Adenylate kinase signature I	PR00094A 9.62 2.43e-09 34-47
786	PR00830	Endopeptidase La (Lon) serine protease (S16) signature I	PR00830A 8.52 4.50e-09 37-56
786	IPB001482	Bacterial type II secretion system protein E	IPB001482B 12.05 4.60e-09 390-412 IPB000135D 2.13 4.73e-09 191-215
786	IPB000850	Adenylate kinase	IPB000850C 18.89 5.03e-09 149-179 IPB000135D 2.13 6.00e-09 190-214
788	PR00452	SH3 domain signature II	PR00452B 11.47 6.03e-09 14-29
789	PR00452	SH3 domain signature II	PR00452B 11.47 6.03e-09 87-102
790	IPB001820	Tissue inhibitors of metalloproteinases	IPB001820C 11.81 1.56e-15 73-85 IPB001820B 10.75 2.44e-14 54-64 IPB001820D 16.18 9.10e-14 91-105 IPB001820A 8.17 2.52e-11 16-29
791	IPB001304	C-type lectin domain	IPB001304A 17.98 3.00e-17 149-173
791	PR01408	Macrophage scavenger receptor signature VI	PR01408F 9.76 4.87e-09 64-88
792	IPB002213	UDP-glucuronosyl and UDP-glucosyl transferase	IPB002213 27.73 3.37e-40 276-322
794	IPB000339	ubiE/COQ5 methyltransferase family	IPB000339D 24.04 6.07e-14 146-188
794	PR00508	S21 class N4 adenine-specific DNA methyltransferase signature II	PR00508B 17.31 3.88e-09 167-187
794	IPB000682	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	IPB000682C 16.46 6.79e-09 68-92
795	PR00237	Rhodopsin-like GPCR superfamily signature III	PR00237C 14.77 1.30e-12 508-530 PR00237B 12.45 8.62e-12 463-484 PR00237D 9.76 3.37e-11 544-565
795	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 2.42e-10 522-533

375

TABLE 3B

795	PR01157	P2 purinoceptor signature IV	PR01157D 16.03 2.98e-09 662-674
795	PR00173	Glutamate-aspartate symporter signature VI	PR00173F 10.23 9.45e-09 705-724 PR00237F 14.34 9.56e-09 645-669
799	PR01539	Interleukin-1 receptor type II precursor signature IX	PR01539I 14.65 9.06e-09 162-185
802	IPB000117	Kappa casein	IPB000117D 10.18 8.71e-09 506-540
805	IPB000171	Bacterial-type phytoene dehydrogenase	IPB000171E 7.19 8.20e-09 29-39
806	IPB001774	Delta serrate ligand	IPB001774D 19.23 5.91e-09 50-96
806	IPB000034	Laminin B	IPB000034C 12.97 7.31e-09 84-102
806	IPB000561	EGF-like domain	IPB000561 4.89 8.07e-09 84-92
807	IPB001774	Delta serrate ligand	IPB001774D 19.23 5.91e-09 50-96
807	IPB000034	Laminin B	IPB000034C 12.97 7.31e-09 84-102
807	IPB000561	EGF-like domain	IPB000561 4.89 8.07e-09 84-92
808	IPB001774	Delta serrate ligand	IPB001774D 19.23 5.91e-09 50-96
808	IPB000034	Laminin B	IPB000034C 12.97 7.31e-09 84-102
808	IPB000561	EGF-like domain	IPB000561 4.89 8.07e-09 84-92
809	PR00436	Interleukin-8 signature I	PR00436A 15.20 9.36e-10 14-37
810	IPB001187	Tissue Factor (TF)	IPB001187G 15.20 7.00e-10 40-76
811	IPB001039	"Major histocompatibility complex protein, Class I"	IPB001039B 27.55 8.79e-09 98-149
812	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 113-150 IPB003006B 20.23 9.14e-12 406-443 IPB003006B 20.23 1.00e-11 213-250
812	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.23e-11 512-535 IPB003006B 20.23 6.40e-10 19-56 IPB003006B 20.23 9.64e-10 505-542 IPB003006B 20.23 8.62e-09 311-348 PR01536C 19.92 9.19e-09 120-143
813	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 428-465 IPB003006B 20.23 8.71e-12 1996-2033 IPB003006B 20.23 9.14e-12 2289-2326 IPB003006B 20.23 1.00e-11 2096-2133
813	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.10e-11 1707-1730 PR01536C 19.92 9.23e-11 2395-2418 IPB003006B 20.23 4.60e-10 1700-1737 IPB003006B 20.23 6.40e-10 1902-1939 IPB003006B 20.23 8.92e-10 1603-1640 IPB003006B 20.23 9.64e-10 2388-2425 IPB003006B 20.23 3.42e-09 1506-1543
813	PR01076	Caldesmon signature IV	PR01076D 8.07 5.07e-09 1457-1478 IPB003006B 20.23 7.58e-09 1799-1836 IPB003006B 20.23 8.62e-09 2194-2231 PR01536C 19.92 9.19e-09 2003-2026
813	PR01472	Intercellular adhesion molecule/vascular cell adhesion molecule-1 signature I	PR01472A 16.78 9.64e-09 1755-1771
814	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 7.60e-16 219-233
814	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 623-660 IPB003006B 20.23 8.71e-12 2191-2228 IPB003006B 20.23 9.14e-12 2484-2521 IPB003006B 20.23 1.00e-11 2291-2328
814	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.10e-11 1902-1925 PR01536C 19.92 9.23e-11 2590-2613 IPB003006B 20.23 4.60e-10 1895-1932 IPB003006B 20.23 6.40e-10 2097-2134

376
TABLE 3B

			IPB003006B 20.23 8.92e-10 1798-1835 IPB003006B 20.23 9.64e-10 2583-2620 IPB003006B 20.23 3.42e-09 1701-1738
814	PR01076	Caldesmon signature IV	PR01076D 8.07 5.07e-09 1652-1673 IPB003006B 20.23 7.58e-09 1994-2031 IPB003006B 20.23 8.62e-09 2389-2426 PR01536C 19.92 9.19e-09 2198-2221
814	PR01472	Intercellular adhesion molecule/vascular cell adhesion molecule-1 signature I	PR01472A 16.78 9.64e-09 1950-1966
816	IPB000074	Apolipoprotein A1/A4/E	IPB000074B 29.17 7.49e-10 117-170 IPB000074B 29.17 8.75e-10 95-148 IPB000074B 29.17 9.20e-10 62-115 IPB000074C 22.23 2.62e-09 90-127 IPB000074C 22.23 4.35e-09 112-149 IPB000074B 29.17 8.48e-09 201-254
817	IPB000074	Apolipoprotein A1/A4/E	IPB000074B 29.17 7.49e-10 117-170 IPB000074B 29.17 8.75e-10 95-148 IPB000074B 29.17 9.20e-10 62-115 IPB000074C 22.23 2.62e-09 90-127 IPB000074C 22.23 4.35e-09 112-149 IPB000074B 29.17 8.48e-09 201-254
819	IPB001211	Phospholipase A2	IPB001211B 17.16 3.12e-31 44-71
819	PR00389	Phospholipase A2 signature III	PR00389C 17.85 2.50e-20 56-74 PR00389B 10.67 6.91e-16 37-55 IPB001211D 11.66 5.50e-14 104-119 PR00389E 13.06 8.20e-14 104-120 IPB001211C 14.62 1.56e-11 79-97
821	IPB001354	Mandelate racemase/muconate lactonizing enzyme family	IPB001354C 32.55 1.00e-24 210-251 IPB001354D 32.92 2.07e-18 281-326 IPB001354B 18.16 3.91e-18 87-113 IPB001354E 9.47 6.23e-09 370-382
822	IPB002164	Nucleosome assembly protein (NAP)	IPB002164B 25.75 1.00e-36 102-138 IPB002164A 24.21 6.40e-34 21-58 IPB002164C 11.48 6.68e-21 151-170
822	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 5.27e-13 285-309 IPB000135D 2.13 1.41e-11 286-310 IPB000135D 2.13 1.82e-11 283-307 IPB000135D 2.13 3.76e-11 289-313 IPB000135D 2.13 3.97e-11 287-311 IPB000135D 2.13 4.27e-11 288-312 IPB002164D 9.19 7.65e-11 232-242 IPB000135D 2.13 1.68e-10 282-306 IPB000135D 2.13 4.03e-10 281-305 IPB000135D 2.13 4.91e-10 284-308
822	IPB001580	Calreticulin family	IPB001580F 2.93 2.35e-09 300-309 IPB000135D 2.13 2.64e-09 280-304 IPB000135D 2.13 6.27e-09 291-315 IPB000135D 2.13 7.27e-09 292-316 IPB000135D 2.13 7.55e-09 279-303 IPB000135D 2.13 8.91e-09 290-314
822	IPB001326	Elongation factor 1 beta/beta'/delta chain	IPB001326C 9.19 9.16e-09 286-301
823	IPB000222	Protein phosphatase 2C subfamily	IPB000222F 19.87 4.94e-15 256-276 IPB000222E 14.28 6.33e-15 228-246 IPB000222G 9.17 1.95e-12 282-295 IPB000222C 6.84 2.08e-12 147-156 IPB000222H 9.33 7.97e-12 318-330

377
TABLE 3B

			IPB000222B 15.80 2.86e-10 115-125 IPB000222D 11.74 2.74e-09 186-203 IPB000222I 8.91 4.72e-09 379-388
824	IPB001007	"von Willebrand factor, type C repeat"	IPB001007B 10.03 1.00e-08 183-192
825	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 59-75
825	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 1-12 PR00245D 9.34 1.53e-13 119-128 PR00245E 8.96 6.81e-12 166-177 PR00245B 13.73 1.00e-10 12-24 IPB000276D 9.40 3.08e-09 165-181
825	PR00237	Rhodopsin-like GPCR superfamily signature V	PR00237E 13.03 3.83e-09 82-105 PR00237G 19.23 1.00e-08 155-181
826	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 173-189
826	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 117-128 PR00245D 9.34 1.53e-13 233-242 PR00245E 8.96 6.81e-12 280-291 PR00245A 10.98 7.14e-12 91-102 PR00245B 13.73 8.14e-10 128-140
826	PR00237	Rhodopsin-like GPCR superfamily signature III	PR00237C 14.77 2.02e-09 103-125 IPB000276D 9.40 3.08e-09 279-295 PR00237E 13.03 3.83e-09 196-219
826	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.17e-09 50-62
826	PR00896	Vasopressin receptor signature II	PR00896B 9.36 7.23e-09 54-65 PR00237G 19.23 1.00e-08 269-295
827	IPB001169	"Integrin beta, C-terminus"	IPB001169J 7.42 4.63e-10 40-53
827	PR01186	Integrin beta subunit signature XI	PR01186K 7.39 7.27e-10 40-53 IPB001169K 27.45 5.50e-09 42-84 PR01186K 7.39 9.75e-09 6-19
828	IPB000198	RhoGAP domain	IPB000198C 16.49 1.28e-10 226-243
829	IPB000859	CUB domain	IPB000859 19.99 7.00e-23 10-45
830	IPB000859	CUB domain	IPB000859 19.99 7.00e-23 10-45
831	PR00193	Myosin heavy chain signature III	PR00193C 11.66 9.77e-24 177-204
831	IPB000857	Core domain in kinesin and myosin motors	IPB000857C 10.82 4.84e-19 175-197 PR00193B 12.36 6.81e-18 125-150 IPB000857D 12.93 8.28e-18 204-242 PR00193A 14.87 8.50e-12 65-84 IPB000857A 15.90 5.58e-11 42-95 IPB000857B 11.35 1.00e-10 106-152
831	PR00364	Disease resistance protein signature I	PR00364A 8.29 4.86e-09 127-142
832	PR00193	Myosin heavy chain signature III	PR00193C 11.66 9.77e-24 177-204
832	IPB000857	Core domain in kinesin and myosin motors	IPB000857C 10.82 4.84e-19 175-197 PR00193B 12.36 6.81e-18 125-150 IPB000857D 12.93 8.28e-18 204-242 IPB000857E 25.07 1.47e-12 288-341 PR00193A 14.87 8.50e-12 65-84 IPB000857A 15.90 5.58e-11 42-95 IPB000857B 11.35 1.00e-10 106-152
832	PR00364	Disease resistance protein signature I	PR00364A 8.29 4.86e-09 127-142 IPB000857F 15.97 6.50e-09 365-397
834	IPB002350	Kazal-type serine protease inhibitor family	IPB002350 31.78 2.86e-18 143-183
834	IPB000716	Thyroglobulin type-1 repeat	IPB000716C 17.62 2.88e-18 336-354 IPB000716D 15.49 7.16e-15 358-372
834	IPB001999	Osteonectin domain	IPB001999E 15.70 7.99e-11 272-318
835	IPB001323	Erythropoietin/thrombopoietin	IPB001323A 17.37 8.31e-10 515-547

378

TABLE 3B

835	PR00251	Bacterial opsin signature I	PR00251A 13.93 9.75e-10 515-534
835	PR00807	Pollen allergen Amb family signature I	PR00807A 16.15 7.41e-09 459-476
836	IPB001323	Erythropoietin/thrombopoietin	IPB001323A 17.37 8.31e-10 515-547
836	PR00251	Bacterial opsin signature I	PR00251A 13.93 9.75e-10 515-534
836	PR00807	Pollen allergen Amb family signature I	PR00807A 16.15 7.41e-09 459-476
838	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 5.50e-13 359-373
838	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 9.33e-10 278-291 PR00019A 11.72 9.33e-10 327-340 PR00019B 11.42 6.73e-09 179-192 PR00019A 11.72 7.27e-09 182-195
840	IPB000243	Proteasome B-type subunit	IPB000243C 13.61 8.80e-09 345-355
841	IPB002889	WSC domain	IPB002889B 11.76 9.36e-11 527-573
841	PR01217	Proline rich extensin signature V	PR01217E 3.04 2.99e-10 534-550 PR01217B 4.82 5.65e-10 533-549 PR01217D 4.57 7.86e-10 529-550
841	IPB000906	ZU5 domain	IPB000906A 22.49 8.91e-10 158-200 PR01217C 4.49 4.80e-09 538-550 IPB000906E 22.11 4.83e-09 162-202 PR01217G 4.02 5.03e-09 529-554
841	PR01415	Ankyrin repeat signature II	PR01415B 10.23 5.88e-09 177-189 PR01415A 12.73 8.00e-09 165-177 PR01415A 12.73 8.75e-09 131-143
841	IPB000925	Pneumovirus attachment glycoprotein G	IPB000925D 14.69 9.33e-09 404-426 PR01217A 5.97 9.62e-09 539-551
842	IPB000416	Outer Capsid protein VP4 (Hemagglutinin)	IPB000416P 15.37 7.40e-09 185-223
843	IPB000416	Outer Capsid protein VP4 (Hemagglutinin)	IPB000416P 15.37 7.00e-09 185-223
844	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006A 17.51 7.11e-09 354-376
845	IPB000998	MAM domain	IPB000998C 18.63 1.95e-12 833-848 IPB000998B 17.20 1.62e-11 761-773
845	PR00020	MAM domain signature I	PR00020A 20.48 3.62e-11 759-777 PR00020C 12.01 8.12e-10 832-843 IPB000998D 18.66 9.61e-10 898-921
845	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006A 17.51 7.11e-09 354-376
845	PR00096	Glutamine amidotransferase superfamily signature III	PR00096C 15.85 9.28e-09 534-547
846	IPB003160	p53-associated protein (MDM2)	IPB003160A 14.23 8.01e-09 82-129
847	IPB002642	Lysophospholipase catalytic domain	IPB002642B 11.84 4.38e-15 1134-1158 IPB002642A 18.37 1.69e-13 1106-1131
847	PR00360	C2 domain signature II	PR00360B 11.64 8.67e-12 839-852 IPB002642G 34.11 6.72e-10 1429-1477
847	IPB000008	C2 domain	IPB000008C 23.37 2.44e-09 812-851
848	IPB002642	Lysophospholipase catalytic domain	IPB002642B 11.84 4.38e-15 383-407 IPB002642A 18.37 1.69e-13 355-380
848	PR00360	C2 domain signature II	PR00360B 11.64 8.67e-12 88-101 IPB002642G 34.11 6.72e-10 678-726 IPB002642E 18.19 6.91e-10 509-534
848	IPB000008	C2 domain	IPB000008C 23.37 2.44e-09 61-100
851	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.43e-13 203-240
851	IPB003531	Short hematopoietin receptor family	IPB003531C 15.87 9.38e-11 449-466

379

TABLE 3B

		1	IPB003006B 20.23 6.54e-09 81-118
852	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.43e-13 199-236
852	IPB003531	Short hematopoietin receptor family 1	IPB003531C 15.87 9.38e-11 445-462 IPB003006B 20.23 6.54e-09 77-114
854	IPB000008	C2 domain	IPB000008C 23.37 7.94e-25 306-345 IPB000008C 23.37 1.17e-16 173-212
854	PR00360	C2 domain signature II	PR00360B 11.64 8.20e-14 200-213 PR00360A 15.18 1.60e-13 304-316
854	PR00399	Synaptotagmin signature II	PR00399B 14.30 1.69e-12 291-304 IPB000008D 14.83 3.45e-11 229-247 IPB000008D 14.83 3.86e-11 361-379 PR00360B 11.64 5.94e-11 333-346 PR00399A 15.05 6.40e-11 145-160 PR00360A 15.18 8.36e-11 173-185 PR00399C 15.89 4.98e-10 348-363 PR00399D 12.72 6.33e-10 368-378 IPB000008C 23.37 9.76e-10 175-214 PR00399B 14.30 6.57e-09 160-173 PR00399A 15.05 8.65e-09 276-291
854	IPB002618	UTP--glucose-1-phosphate uridylyltransferase	IPB002618D 29.24 9.88e-09 182-224
855	IPB002870	Reprolysins family propeptide	IPB002870B 24.73 3.78e-14 141-179 IPB002870E 11.90 4.67e-14 391-403 IPB002870F 18.81 7.00e-13 432-456 IPB002870D 16.31 6.62e-12 360-375
855	IPB001762	Disintegrin	IPB001762A 23.93 1.40e-11 336-376
855	IPB000130	"Neutral zinc metalloproteinases, zinc-binding region"	IPB000130 5.86 5.15e-11 389-399
855	PR00480	Astacin family signature II	PR00480B 14.35 4.54e-10 384-402
855	PR01303	Plasmodium circumsporozoite protein signature IV	PR01303D 10.57 4.71e-10 953-970 PR01303D 10.57 2.75e-09 833-850
855	IPB001670	Iron-containing alcohol dehydrogenase	IPB001670D 13.90 5.50e-09 157-172 IPB002870C 11.01 5.68e-09 317-327 PR01303D 10.57 6.38e-09 552-569
855	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862A 12.54 6.66e-09 540-555
856	IPB003952	Fumarate reductase / succinate dehydrogenase FAD-binding site	IPB003952A 6.70 8.00e-09 14-28
857	PR00833	Pollen allergen Poa pI signature VIII	PR00833H 2.61 4.11e-09 58-72
857	IPB002989	Mycobacterial pentapeptide repeats	IPB002989C 13.82 8.67e-09 48-87
858	PR00833	Pollen allergen Poa pI signature VIII	PR00833H 2.61 4.11e-09 51-65
859	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 8.26e-26 254-306
859	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 6.77e-24 265-318 IPB000885B 19.15 9.30e-24 247-300 IPB000885B 19.15 1.42e-23 244-297 IPB001442A 26.12 5.96e-23 257-309 IPB001442A 26.12 8.83e-23 266-318 IPB001442A 26.12 8.96e-23 239-291 IPB000885B 19.15 9.45
859	PR01408	Macrophage scavenger receptor signature VIII	PR01408H 14.32 5.76e-16 227-246
859	PR00258	Speract receptor signature I	PR00258A 13.56 6.32e-16 333-349 IPB001442A 26.12 8.12e-16 272-324 IPB000885A 11.46 4.16e-15 255-292 IPB000885B 19.15 5.76e-15 274-327

380
TABLE 3B

			IPB000885A 11.46 5.86e-15 270-307 IPB001442A 26.12 7.88e-15 230-282 IPB000885A 11.46 2.87e-14 276-313 IPB000885B 19.15 3.43e-14 229-282 IPB000885B 19.15 4.13e-14 277-330 IPB000885A 11.46 5.44e-14 243-280 IPB000885A 11.46 7.78e-14 285-322 IPB000885B 19.15 7.88e-14 280-333
859	IPB001073	Complement C1q protein	IPB001073A 22.14 8.40e-14 263-297 IPB000885B 19.15 5.21e-13 226-279 IPB001073A 22.14 5.79e-13 269-303 PR00258B 7.94 8.42e-13 352-363 IPB001442B 12.38 9.00e-13 270-290 IPB001442A 26.12 9.16e-13 227-279 IPB001073A 22.14 1.54e-1
859	IPB000817	Prion protein	IPB000817A 8.34 5.85e-10 244-286 IPB001073A 22.14 6.80e-10 287-321 IPB000817A 8.34 8.22e-10 247-289 IPB001442B 12.38 8.46e-10 246-266 IPB000885A 11.46 9.32e-10 234-271 IPB001442A 26.12 9.42e-10 284-336 IPB000885A 11.46 9.61e-10 288-325 IPB001442B 12.38 1.24e-09 264-284 IPB001442A 26.12 1.63e-09 221-273 IPB001073A 22.14 2.83e-09 251-285 IPB001073A 22.14 3.53e-09 284-318 IPB001442B 12.38 4.65e-09 291-311 IPB001442B 12.38 4.77e-09 249-269 IPB001073A 22.14 5.64e-09 278-312 IPB000885A 11.46 5.87e-09 291-328 IPB001442B 12.38 6.11e-09 273-293 IPB001442B 12.38 6.84e-09 294-314 IPB001073A 22.14 7.61e-09 239-273
860	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 8.26e-26 314-366
860	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 4.52e-24 307-360 IPB000885B 19.15 6.77e-24 325-378 IPB000885B 19.15 1.69e-23 304-357 IPB001442A 26.12 5.96e-23 317-369 IPB001442A 26.12 6.35e-23 299-351 IPB001442A 26.12 8.83e-23 326-378 IPB000885B 19.15 1.26
860	PR01408	Macrophage scavenger receptor signature VIII	PR01408H 14.32 5.76e-16 287-306
860	PR00258	Speract receptor signature I	PR00258A 13.56 6.32e-16 393-409 IPB001442A 26.12 8.12e-16 332-384 IPB000885A 11.46 4.16e-15 315-352 IPB000885B 19.15 5.76e-15 334-387 IPB000885A 11.46 5.86e-15 330-367 IPB000885B 19.15 7.35e-15 289-342 IPB001442A 26.12 7.88e-15 290-342 IPB000885A 11.46 2.87e-14 336-373 IPB000885B 19.15 4.13e-14 337-390 IPB000885A 11.46 5.91e-14 303-340 IPB000885A 11.46 7.78e-14 345-382 IPB000885B 19.15 7.88e-14 340-393
860	IPB001073	Complement C1q protein	IPB001073A 22.14 8.40e-14 323-357 IPB000885B 19.15 5.70e-13 286-339

381

TABLE 3B

			IPB001073A 22.14 5.79e-13 329-363 IPB001442A 26.12 7.28e-13 287-339 PR00258B 7.94 8.42e-13 412-423 IPB001442B 12.38 9.00e-13 330-350 IPB001073A 22.14 1.54e-1
860	IPB000817	Prion protein	IPB000817A 8.34 5.65e-10 304-346 IPB001073A 22.14 6.03e-10 311-345 IPB001073A 22.14 6.80e-10 347-381 PR00258C 9.05 7.15e-10 427-437 PR00258D 14.29 8.06e-10 458-472 IPB000817A 8.34 8.42e-10 307-349 IPB001442A 26.12 9.42e-10 344-396 IPB000885A 11.46 9.61e-10 348-385 IPB001073A 22.14 9.69e-10 299-333 IPB000885B 19.15 9.83e-10 283-336 IPB000885A 11.46 9.90e-10 294-331 IPB001442B 12.38 1.24e-09 324-344 IPB001442A 26.12 2.41e-09 281-333 IPB001442B 12.38 2.70e-09 309-329 IPB001073A 22.14 3.53e-09 344-378 IPB001442B 12.38 4.65e-09 351-371 IPB001073A 22.14 5.64e-09 338-372 IPB000885A 11.46 5.87e-09 351-388 IPB001442B 12.38 6.11e-09 333-353 IPB001442B 12.38 6.84e-09 354-374 PR01408B 9.21 9.84e-09 58-83
862	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 8.20e-22 222-247 IPB000822 14.67 5.09e-21 306-331 IPB000822 14.67 5.50e-20 474-499 IPB000822 14.67 7.00e-20 446-471 IPB000822 14.67 3.25e-19 390-415 IPB000822 14.67 4.00e-19 194-219 IPB000822 14.67 7.00e-19 278-303 IPB000822 14.67 4.46e-18 362-387 IPB000822 14.67 6.14e-17 250-275 IPB000822 14.67 3.40e-16 418-443 IPB000822 14.67 4.00e-16 334-359
862	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 5.85e-14 415-428 PR00048A 9.94 8.07e-13 219-232 PR00048A 9.94 3.12e-12 387-400 PR00048A 9.94 4.71e-12 247-260 PR00048A 9.94 4.71e-12 331-344 PR00048B 5.52 7.00e-12 487-496
862	IPB001275	DM DNA binding domain	IPB001275 19.17 7.04e-12 266-305 PR00048A 9.94 7.88e-12 499-512 PR00048A 9.94 1.95e-11 471-484 PR00048A 9.94 4.32e-11 443-456 PR00048B 5.52 5.50e-11 319-328 PR00048A 9.94 1.00e-10 191-204 IPB001275 19.17 1.36e-10 294-333 IPB001275 19.17 1.49e-10 350-389 PR00048A 9.94 5.09e-10 303-316 IPB001275 19.17 5.14e-10 378-417
862	IPB002817	ThiC family	IPB002817H 11.39 5.42e-10 217-232 PR00048A 9.94 5.91e-10 359-372 IPB001275 19.17 8.18e-10 182-221 IPB001275 19.17 9.15e-10 322-361 PR00048B 5.52 9.36e-10 375-384

382
TABLE 3B

			IPB001275 19.17 9.39e-10 210-249 IPB001275 19.17 9.39e-10 238-277 PR00048B 5.52 2.00e-09 207-216 IPB000822 14.67 2.13e-09 502-527 PR00048B 5.52 2.50e-09 459-468 IPB001275 19.17 2.71e-09 462-501 PR00048B 5.52 3.00e-09 403-412 IPB001275 19.17 3.62e-09 406-445 PR00048A 9.94 4.38e-09 275-288
862	IPB000306	"FYVE Zn-finger, rabphilin/VPS27/FAB1 type"	IPB000306 8.96 4.71e-09 218-230 PR00048B 5.52 5.50e-09 291-300 IPB000306 8.96 5.76e-09 498-510 IPB000306 8.96 6.03e-09 302-314 PR00048B 5.52 7.00e-09 235-244 IPB002817H 11.39 7.34e-09 301-316 IPB001275 19.17 8.18e-09 434-473
862	IPB002634	BolA-like protein	IPB002634A 23.30 8.62e-09 243-277
864	IPB000571	Zinc finger C-x8-C-x5-C-x3-H type	IPB000571 11.41 6.54e-10 66-76
864	PR01218	Pistil-specific extensin-like signature II	PR01218B 8.47 9.12e-09 140-163
865	PR00320	G protein beta WD-40 repeat signature II	PR00320B 12.82 5.68e-10 225-239 PR00320A 13.15 7.48e-10 225-239
865	IPB001680	G-protein beta WD-40 repeats	IPB001680 10.43 4.15e-09 227-238 PR00320C 12.32 9.67e-09 225-239
867	IPB000954	Aminotransferase class-III pyridoxal- phosphate	IPB000954B 21.02 9.25e-25 291-330 IPB000954A 20.25 7.12e-18 98-127 IPB000954D 13.61 5.74e-17 377-395 IPB000954C 12.88 9.44e-14 340-355
868	IPB000954	Aminotransferase class-III pyridoxal- phosphate	IPB000954B 21.02 9.25e-25 188-227 IPB000954D 13.61 5.74e-17 274-292 IPB000954C 12.88 9.44e-14 237-252
869	IPB001254	"Serine proteases, trypsin family"	IPB001254C 16.54 2.50e-17 270-289
869	IPB000177	Apple domain	IPB000177O 14.39 1.11e-15 267-295 IPB001254A 9.98 6.14e-15 88-104
869	PR00722	Chymotrypsin serine protease family (S1) signature III	PR00722C 10.74 3.08e-14 236-248 PR00722A 12.06 4.54e-14 89-104 IPB001254B 15.01 7.14e-14 237-260
869	IPB000001	Kringle	IPB000001D 11.31 7.56e-12 88-104 IPB000001H 12.24 2.50e-11 239-249 IPB000177N 10.17 3.23e-11 229-263 IPB000177K 13.19 2.57e-10 90-122 PR00722B 12.69 6.85e-10 145-159
873	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862F 29.39 6.19e-15 343-390
873	PR00010	Type II EGF-like signature I	PR00010A 12.91 4.94e-13 46-57
873	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 7.55e-13 541-556 IPB001862F 29.39 8.07e-13 553-600 IPB001862F 29.39 9.14e-13 515-562 IPB001862F 29.39 3.07e-12 35-82 IPB001862F 29.39 3.79e-12 73-120 IPB001862F 29.39 4.10e-12 304-351 IPB000152 8.86 6.04e-12 61-76 IPB001862F 29.39 8.45e-12 477-524 IPB001862F 29.39 8.45e-12 1031-1078 IPB000152 8.86 3.89e-11 137-152 IPB001862F 29.39 4.00e-11 153-200 IPB000152 8.86 4.86e-11 179-194

383

TABLE 3B

			IPB001862F 29.39 6.70e-11 381-428 PR00010C 6.98 7.38e-11 374-384
873	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 7.63e-11 137-148 PR00010C 6.98 9.25e-11 66-76 IPB001862F 29.39 9.50e-11 265-312 PR00010A 12.91 1.00e-10 564-575 IPB000152 8.86 1.84e-10 369-384 PR00010A 12.91 2.38e-10 354-365 IPB001862F 29.39 2.63e-10 111-158 PR00010A 12.91 2.73e-10 488-499
873	PR00764	Complement C9 signature VI	PR00764F 15.74 2.92e-10 170-190
873	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 3.03e-10 374-384 PR00764F 15.74 3.16e-10 52-72 PR00764F 15.74 3.52e-10 321-341 PR00010C 6.98 3.90e-10 546-556 IPB001881B 12.28 4.00e-10 541-552 IPB000152 8.86 4.66e-10 503-518 IPB001881A 8.72 4.86e-10 280-289 PR00010A 12.91 5.50e-10 122-133
873	IPB002899	EB module	IPB002899B 11.81 5.59e-10 243-255 IPB000152 8.86 6.06e-10 407-422 IPB000033B 7.05 6.23e-10 296-306 IPB000152 8.86 6.63e-10 291-306 IPB001881A 8.72 7.43e-10 319-328 IPB001881A 8.72 7.43e-10 530-539 IPB001881A 8.72 8.07e-10 126-135 IPB001881B 12.28 8.29e-10 255-266 IPB000152 8.86 8.31e-10 23-38 PR00764F 15.74 8.44e-10 360-380 PR00764F 15.74 8.44e-10 570-590 IPB001881A 8.72 9.36e-10 168-177 PR00764F 15.74 9.52e-10 398-418 IPB000152 8.86 9.72e-10 255-270 PR00010C 6.98 1.00e-09 296-306 IPB001881A 8.72 2.20e-09 1046-1055
873	PR00011	Type III EGF-like signature II	PR00011B 13.08 2.23e-09 63-81 IPB001881B 12.28 2.57e-09 179-190 IPB001881A 8.72 2.80e-09 358-367
873	IPB003884	Factor I membrane attack complex	IPB003884C 13.00 2.83e-09 572-590
873	IPB000561	EGF-like domain	IPB000561 4.89 2.93e-09 626-634 PR00010C 6.98 3.63e-09 28-38 IPB000561 4.89 4.21e-09 378-386
873	IPB000359	Cystine-knot domain	IPB000359A 23.24 4.33e-09 70-94 IPB000561 4.89 4.86e-09 108-116 IPB000359A 23.24 4.91e-09 108-132 PR00010C 6.98 6.05e-09 184-194 IPB001881A 8.72 6.40e-09 50-59
873	IPB000034	Laminin B	IPB000034C 12.97 6.49e-09 70-88 PR00010A 12.91 7.27e-09 164-175 PR00010A 12.91 7.27e-09 315-326
873	IPB001886	Laminin N-terminal (Domain VI)	IPB001886C 24.54 7.40e-09 300-339 IPB000561 4.89 7.43e-09 223-231 IPB000561 4.89 7.43e-09 550-558 PR00011D 12.12 7.81e-09 371-389 IPB000152 8.86 8.11e-09 330-345 IPB000359A 23.24 8.24e-09 512-536
873	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 8.62e-09 75-120 PR00010C 6.98 8.68e-09 412-422

384

TABLE 3B

			PR00764F 15.74 9.20e-09 282-302 IPB000033B 7.05 9.29e-09 546-556 IPB001862F 29.39 9.36e-09 591-638 IPB001881A 8.72 9.40e-09 568-577 PR00764F 15.74 9.43e-09 532-552 PR00010A 12.91 9.45e-09 526-537 PR00011D 12.12 9.74e-09 25-43 IPB000033B 7.05 1.00e-08 66-76
874	PR00960	LmbP protein signature I	PR00960A 10.63 4.67e-09 78-93
875	IPB000043	S-adenosyl-L-homocysteine hydrolase	IPB000043D 24.21 1.00e-40 235-289 IPB000043E 21.11 1.00e-40 298-350 IPB000043A 16.26 4.72e-33 119-156 IPB000043H 17.16 1.72e-29 459-493 IPB000043F 16.20 2.55e-24 351-377 IPB000043G 18.51 3.25e-24 411-448 IPB000043B 18.62 5.95e-23 158-191 IPB000043G 18.51 7.16e-15 412-449 IPB000043C 8.96 9.61e-15 202-216
878	IPB002181	Fibrinogen beta and gamma chains C-terminal globular domain	IPB002181B 20.16 7.49e-24 181-217 IPB002181D 29.18 7.32e-15 243-283 IPB002181C 15.87 2.64e-10 222-234
879	IPB002181	Fibrinogen beta and gamma chains C-terminal globular domain	IPB002181B 20.16 7.49e-24 181-217 IPB002181D 29.18 7.32e-15 243-283 IPB002181C 15.87 2.64e-10 222-234
880	IPB002181	Fibrinogen beta and gamma chains C-terminal globular domain	IPB002181B 20.16 7.49e-24 181-217 IPB002181D 29.18 7.32e-15 243-283 IPB002181C 15.87 2.64e-10 222-234
883	IPB002027	Amino acid permease	IPB002027D 22.00 4.13e-25 325-364 IPB002027C 19.67 2.74e-22 244-282 IPB002027A 18.88 3.77e-16 47-75 IPB002027B 12.67 7.97e-12 180-199
884	IPB001772	Kinase associated domain 1	IPB001772E 24.88 4.03e-10 620-659
884	IPB000861	PKN/rhopilin/rhotekin rho-binding repeat	IPB000861D 13.61 7.34e-10 97-133
884	IPB000961	Protein kinase C-terminal domain	IPB000961A 16.82 8.45e-09 99-133
884	IPB003527	MAP kinase	IPB003527D 21.53 9.15e-09 462-503
885	IPB001304	C-type lectin domain	IPB001304A 17.98 8.04e-14 34-58
891	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.72e-10 103-140
891	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 1.31e-09 155-169 PR00049D 0.00 6.80e-09 156-170
892	PR00503	Bromodomain signature IV	PR00503D 19.24 3.57e-21 421-440
892	IPB001487	Bromodomain	IPB001487B 17.44 2.13e-19 412-433 PR00503B 10.44 4.37e-19 94-110 IPB001487A 11.44 5.20e-19 95-113 PR00503C 19.09 4.00e-17 110-128 IPB001487A 11.44 9.53e-16 388-406 PR00503A 14.57 4.00e-14 78-91 PR00503B 10.44 8.64e-14 387-403
892	IPB001359	Synapsin	IPB001359H 22.58 1.65e-13 752-802 PR00503D 19.24 9.25e-13 128-147 IPB001487B 17.44 1.58e-12 119-140 PR00503C 19.09 6.70e-11 403-421
892	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 8.87e-11 755-769 PR00049D 0.00 9.47e-11 756-770 IPB001359H 22.58 9.70e-11 979-1029
892	PR00209	Alpha/beta gliadin family signature II	PR00209B 4.73 4.80e-10 966-984

385

TABLE 3B

892	IPB003861	E4 protein	IPB003861B 9.06 4.86e-10 979-993
892	IPB001505	"Cu(A) centre of cytochrome c oxidase, subunit II and nitrous oxide reductase"	IPB001505B 15.93 5.94e-10 406-455 PR00209B 4.73 6.90e-10 968-986 IPB001359H 22.58 7.40e-10 753-803
892	PR01471	Histamine H3 receptor signature V	PR01471E 5.41 7.44e-10 765-780 IPB001359H 22.58 7.77e-10 962-1012 PR00209B 4.73 9.80e-10 752-770 IPB001505A 18.04 1.17e-09 93-140 PR00049D 0.00 2.22e-09 748-762 IPB003861B 9.06 3.15e-09 763-777 PR00049D 0.00 3.29e-09 972-986 IPB001359H 22.58 3.88e-09 757-807 PR01471E 5.41 4.03e-09 981-996 PR01471E 5.41 4.23e-09 1019-1034 IPB003861B 9.06 4.52e-09 754-768
892	IPB003351	Dishevelled specific domain	IPB003351C 13.82 5.13e-09 485-524 IPB001359H 22.58 5.19e-09 941-991 PR01471E 5.41 5.99e-09 755-770 PR00503A 14.57 6.81e-09 371-384 IPB001359H 22.58 7.03e-09 765-815 IPB001359H 22.58 7.03e-09 970-1020
892	PR01217	Proline rich extensin signature IV	PR01217D 4.57 7.49e-09 239-260
892	PR01503	Treacher Collins syndrome protein Treacle signature II	PR01503B 3.77 7.64e-09 702-715
892	IPB000574	Tymovirus coat protein	IPB000574A 32.18 7.78e-09 254-301
892	PR00910	Luteovirus ORF6 protein signature I	PR00910A 2.74 8.07e-09 255-267 IPB001359H 22.58 8.25e-09 978-1028 IPB001359H 22.58 8.51e-09 193-243 IPB001359H 22.58 8.51e-09 745-795 IPB001359H 22.58 9.04e-09 754-804
892	IPB001978	Troponin	IPB001978B 22.99 9.15e-09 530-561 PR00209B 4.73 9.90e-09 758-776
893	IPB003112	Olfactomedin-like domain	IPB003112C 13.54 4.69e-33 343-383 IPB003112E 16.12 5.24e-33 416-458 IPB003112B 14.91 6.65e-27 269-320 IPB003112D 17.44 9.58e-23 384-410 IPB003112A 14.44 2.97e-13 230-245
893	PR01444	Latrophilin receptor signature V	PR01444E 11.17 7.70e-12 346-361
893	PR00952	Type III secretion system inner membrane Q protein family signature III	PR00952C 21.25 2.04e-09 7-29
893	IPB002862	Protein of unknown function DUF16	IPB002862C 11.30 9.59e-09 80-102
894	IPB002350	Kazal-type serine protease inhibitor family	IPB002350 31.78 4.12e-21 92-132
894	PR00290	Kazal-type serine protease inhibitor signature I	PR00290A 13.80 3.61e-12 92-102
894	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.36e-10 390-427
894	PR00450	Recoverin family signature III	PR00450C 11.99 5.04e-09 182-203
895	IPB001511	Aminotransferases class-I	IPB001511B 11.54 3.14e-11 177-191
895	PR00753	1-aminocyclopropane-1-carboxylate synthase signature V	PR00753E 10.09 9.22e-11 171-195 IPB001511C 12.45 9.07e-10 243-256
896	IPB001781	LIM domain	IPB001781 11.42 3.37e-12 102-112 IPB001781 11.42 2.04e-10 173-183 IPB001781 11.42 4.60e-09 43-53 IPB001781 11.42 7.90e-09 231-241
896	IPB003452	Stem cell factor	IPB003452C 13.68 9.29e-09 525-558

386

TABLE 3B

897	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 5.29e-29 512-553
897	IPB001772	Kinase associated domain 1	IPB001772B 18.27 4.79e-24 409-454
897	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 2.80e-19 516-554
897	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861G 13.73 9.60e-16 518-567
897	IPB000959	POLO box duplicated region	IPB000959C 23.49 8.03e-15 491-543
897	IPB003527	MAP kinase	IPB003527G 17.26 8.94e-15 586-623 IPB001772E 24.88 2.25e-14 574-613 IPB000961B 17.79 2.37e-14 412-443 IPB001245A 22.45 6.88e-14 460-500 IPB000961A 16.82 7.75e-14 355-389 IPB001772C 20.66 9.62e-14 455-485 IPB001772D 21.67 4.73e-13 523-562 IPB000959B 15.68 3.18e-11 444-484 IPB001772A 13.64 5.22e-11 353-384 IPB003527D 21.53 6.02e-11 509-550 IPB000861E 16.40 9.36e-11 399-444
897	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 9.65e-10 520-574 IPB003527C 14.70 2.54e-09 452-500 IPB000861D 13.61 2.99e-09 353-389 IPB000961C 15.48 3.45e-09 467-501
897	PR00109	Tyrosine kinase catalytic domain signature II	PR00109B 11.07 3.81e-09 467-485 IPB000959D 27.01 5.31e-09 567-619 IPB000959A 7.12 7.62e-09 356-368
898	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 5.29e-29 699-740
898	IPB001772	Kinase associated domain 1	IPB001772B 18.27 4.79e-24 596-641
898	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 2.80e-19 703-741
898	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861G 13.73 9.60e-16 705-754
898	IPB000959	POLO box duplicated region	IPB000959C 23.49 8.03e-15 678-730
898	IPB003527	MAP kinase	IPB003527G 17.26 8.94e-15 773-810 IPB001772E 24.88 2.25e-14 761-800 IPB000961B 17.79 2.37e-14 599-630 IPB001245A 22.45 6.88e-14 647-687 IPB000961A 16.82 7.75e-14 542-576 IPB001772C 20.66 9.62e-14 642-672 IPB001772D 21.67 4.73e-13 710-749
898	IPB003533	Doublecortin	IPB003533F 11.80 5.30e-12 161-194 IPB000959B 15.68 3.18e-11 631-671 IPB001772A 13.64 5.22e-11 540-571 IPB003527D 21.53 6.02e-11 696-737 IPB000861E 16.40 9.36e-11 586-631
898	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 9.65e-10 707-761 IPB003527C 14.70 2.54e-09 639-687 IPB000861D 13.61 2.99e-09 540-576 IPB000961C 15.48 3.45e-09 654-688
898	PR00109	Tyrosine kinase catalytic domain signature II	PR00109B 11.07 3.81e-09 654-672 IPB000959D 27.01 5.31e-09 754-806 IPB000959A 7.12 7.62e-09 543-555 IPB003533E 7.28 8.25e-09 105-144
900	IPB001073	Complement C1q protein	IPB001073B 20.88 6.00e-26 131-165 IPB001073A 22.14 4.48e-20 85-119
900	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 9.63e-20 54-107
900	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 4.27e-19 55-107 IPB000885B 19.15 7.48e-19 60-113 IPB000885A 11.46 1.97e-18 62-99 IPB000885A 11.46 2.94e-18 68-105

387

TABLE 3B

900	PR00007	Complement C1Q domain signature III	PR00007C 16.13 3.67e-18 199-220 IPB001442A 26.12 1.11e-17 64-116 PR00007A 20.64 1.84e-17 124-150 IPB001442A 26.12 1.87e-17 70-122 IPB000885B 19.15 5.39e-17 57-110 IPB000885A 11.46 6.96e-17 65-102 IPB000885B 19.15 8.87e-17 51-
900	IPB000817	Prion protein	IPB000817A 8.34 3.27e-09 51-93 IPB000885A 11.46 3.66e-09 19-56 IPB001442A 26.12 4.13e-09 12-64 IPB000885B 19.15 4.19e-09 26-79 IPB000885A 11.46 4.77e-09 86-123 IPB001442A 26.12 4.83e-09 24-76 IPB001442B 12.38 5.99e-09 37-57 IPB001442A 26.12 6.17e-09 21-73 IPB000885B 19.15 7.55e-09 36-89 IPB001442B 12.38 7.57e-09 71-91 IPB001442A 26.12 8.36e-09 9-61 IPB001442B 12.38 8.54e-09 89-109 IPB001073A 22.14 8.59e-09 30-64 IPB000885B 19.15 8.69e-09 78-131 IPB001442B 12.38 9.64e-09 74-94
901	IPB000074	Apolipoprotein A1/A4/E	IPB000074A 11.45 9.84e-09 7-24
902	IPB002360	Involucrin	IPB002360C 15.36 3.06e-14 407-448
902	PR00209	Alpha/beta gliadin family signature II	PR00209B 4.73 5.94e-12 427-445
902	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 8.67e-11 183-207 IPB000135D 2.13 2.96e-10 184-208
902	IPB001580	Calreticulin family	IPB001580F 2.93 4.94e-10 189-198 IPB001580F 2.93 4.94e-10 190-199 IPB001580F 2.93 4.94e-10 191-200 IPB002360C 15.36 5.93e-10 416-457 IPB000135D 2.13 7.46e-10 186-210 IPB000135D 2.13 7.46e-10 187-211 IPB000135D 2.13 9.22e-10 185-209 IPB002360C 15.36 2.50e-09 396-437 IPB002360C 15.36 2.50e-09 415-456 IPB000135D 2.13 3.55e-09 182-206 IPB000135D 2.13 4.27e-09 188-212 IPB000135D 2.13 4.91e-09 181-205
902	IPB001359	Synapsin	IPB001359H 22.58 5.19e-09 421-471 IPB002360C 15.36 5.20e-09 404-445
902	IPB001422	Neuromodulin (GAP-43)	IPB001422C 16.82 5.61e-09 184-219 IPB002360C 15.36 5.70e-09 413-454 IPB002360C 15.36 6.10e-09 389-430
902	IPB003753	"Exonuclease VII, large subunit"	IPB003753F 28.29 7.54e-09 382-432 IPB002360C 15.36 8.80e-09 419-460
905	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 5.50e-13 37-51
905	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 9.33e-10 5-18
906	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.83e-11 55-92
908	PR00457	Animal haem peroxidase signature V	PR00457E 19.97 8.45e-24 1041-1067 PR00457D 18.35 1.53e-20 1016-1036 PR00457C 18.81 9.42e-15 998-1016 PR00457G 14.17 4.48e-14 1221-1241 PR00457H 14.82 5.85e-13 1292-1306 PR00457F 14.42 6.32e-12 1094-1104

TABLE 3B

908	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 1.00e-10 156-170 PR00457B 12.43 2.29e-10 846-861
908	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 2.80e-10 352-389 IPB003006B 20.23 8.92e-10 448-485 IPB003006B 20.23 9.28e-10 259-296
909	PR00457	Animal haem peroxidase signature V	PR00457E 19.97 8.45e-24 1072-1098 PR00457D 18.35 1.53e-20 1047-1067 PR00457C 18.81 9.42e-15 1029-1047 PR00457G 14.17 4.48e-14 1252-1272 PR00457H 14.82 5.85e-13 1323-1337 PR00457F 14.42 6.32e-12 1125-1135
909	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 1.00e-10 187-201 PR00457B 12.43 2.29e-10 877-892
909	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 2.80e-10 383-420 IPB003006B 20.23 8.92e-10 479-516 IPB003006B 20.23 9.28e-10 290-327
910	PR00457	Animal haem peroxidase signature V	PR00457E 19.97 8.45e-24 934-960 PR00457D 18.35 1.53e-20 909-929 PR00457C 18.81 9.42e-15 891-909 PR00457G 14.17 4.48e-14 1114-1134 PR00457H 14.82 5.85e-13 1185-1199 PR00457F 14.42 6.32e-12 987-997 PR00457B 12.43 2.29e-10 739-754
910	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 2.80e-10 329-366 IPB003006B 20.23 8.92e-10 425-462 IPB003006B 20.23 9.28e-10 236-273
910	PR00019	Leucine-rich repeat signature II	PR00019B 11.42 6.73e-09 73-86
911	PR00010	Type II EGF-like signature I	PR00010A 12.91 7.75e-13 43-54
911	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862F 29.39 5.45e-12 925-972 IPB001862F 29.39 7.21e-12 559-606
911	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 7.48e-12 117-132 PR00010A 12.91 1.00e-11 102-113 PR00010A 12.91 4.27e-11 168-179
911	IPB000561	EGF-like domain	IPB000561 4.89 8.00e-11 88-96 IPB001862F 29.39 8.70e-11 846-893
911	IPB000359	Cystine-knot domain	IPB000359A 23.24 2.71e-10 843-867 PR00010C 6.98 3.61e-10 122-132
911	IPB001881	Calcium-binding EGF-like domain	IPB001881A 8.72 4.21e-10 574-583 IPB000152 8.86 4.66e-10 183-198 IPB000561 4.89 4.75e-10 126-134 IPB001881A 8.72 6.79e-10 940-949 PR00010C 6.98 7.10e-10 877-887 PR00010C 6.98 7.68e-10 230-240 PR00010C 6.98 1.22e-09 590-600
911	PR00764	Complement C9 signature VI	PR00764F 15.74 1.34e-09 942-962 IPB001881B 12.28 1.78e-09 183-194 PR00764F 15.74 3.28e-09 576-596 IPB000359A 23.24 3.35e-09 88-112 IPB000561 4.89 4.21e-09 881-889 IPB000152 8.86 4.79e-09 834-849 IPB001862F 29.39 6.23e-09 91-138 IPB001881A 8.72 7.00e-09 106-115 IPB001881B 12.28 7.65e-09 117-128 PR00010C 6.98 7.80e-09 84-94
911	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 8.34e-09 877-887 IPB000152 8.86 8.58e-09 872-887
911	IPB003884	Factor I membrane attack complex	IPB003884F 16.26 8.71e-09 177-192

TABLE 3B

913	IPB002151	Kinesin light chain repeat	IPB002151B 14.23 8.01e-10 240-292
913	IPB000421	Coagulation factor 5/8 type C domain (FA58C)	IPB000421A 21.21 7.85e-09 43-62
913	IPB002360	Involucrin	IPB002360C 15.36 8.00e-09 373-414
914	IPB003117	Regulatory subunit of type II PKA R-subunit	IPB003117C 17.01 1.00e-40 147-187 IPB003117D 18.87 1.00e-40 198-238 IPB003117G 17.45 8.50e-33 341-375 IPB003117A 22.23 5.50e-26 24-56 IPB003117E 18.84 5.85e-23 287-315
914	IPB000595	Cyclic nucleotide-binding domain	IPB000595C 23.31 6.82e-21 321-346
914	PR00103	cAMP-dependent protein kinase signature II	PR00103B 10.32 7.00e-18 173-187 IPB000595B 15.72 7.50e-18 279-302 IPB003117F 17.26 1.00e-17 323-337 IPB000595B 15.72 4.43e-16 161-184 PR00103A 9.07 7.75e-16 158-172 IPB003117C 17.01 2.96e-15 265-305 IPB003117D 18.87 4.14e-15 322-362 PR00103E 12.91 5.91e-14 355-367 PR00103D 10.18 2.93e-13 334-345 IPB000595C 23.31 4.60e-13 197-222 PR00103C 13.28 1.84e-11 322-331 PR00103D 10.18 2.98e-10 210-221 IPB003117E 18.84 3.57e-10 157-185 IPB003117E 18.84 5.43e-10 275-303 IPB003117F 17.26 1.50e-09 199-213 PR00103A 9.07 8.11e-09 276-290
915	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478B 6.12 4.94e-09 602-611
916	IPB000907	Lipoxygenase	IPB000907J 20.31 5.50e-37 521-563 IPB000907G 22.23 1.87e-34 371-413 IPB000907F 21.29 1.00e-28 338-370 IPB000907I 27.52 9.79e-28 460-513
916	PR00467	Mammalian lipoxygenase signature VI	PR00467F 12.25 9.41e-22 418-440
916	PR00087	Lipoxygenase signature III	PR00087C 13.32 1.39e-21 373-393 IPB000907C 16.09 7.17e-21 221-247 IPB000907E 15.16 1.00e-18 296-320 PR00467E 9.17 2.10e-17 293-312 PR00467D 17.16 9.57e-17 196-217 IPB000907D 18.70 2.67e-16 262-289 PR00087A 20.06 3.52e-15 335-352 PR00087B 13.69 5.11e-15 353-370 IPB000907B 14.10 2.50e-13 160-175 PR00467A 8.38 3.29e-13 11-28 IPB000907H 18.37 5.86e-13 434-450 PR00467B 14.98 5.88e-12 57-76 PR00467G 16.61 3.37e-11 576-593 IPB000907A 16.20 4.21e-10 94-103 PR00467C 9.34 7.65e-10 134-148
917	IPB000907	Lipoxygenase	IPB000907C 16.09 7.17e-21 194-220 IPB000907E 15.16 1.00e-18 269-293
917	PR00467	Mammalian lipoxygenase signature V	PR00467E 9.17 2.10e-17 266-285 PR00467D 17.16 9.57e-17 169-190 IPB000907D 18.70 2.67e-16 235-262 IPB000907B 14.10 2.50e-13 131-146 PR00467A 8.38 3.29e-13 11-28 PR00467B 14.98 5.88e-12 57-76 IPB000907A 16.20 4.21e-10 94-103

390

TABLE 3B

918	IPB000907	Lipoxygenase	IPB000907C 16.09 7.17e-21 223-249 IPB000907E 15.16 1.00e-18 298-322
918	PR00467	Mammalian lipoxygenase signature V	PR00467E 9.17 2.10e-17 295-314 PR00467D 17.16 9.57e-17 198-219 IPB000907D 18.70 2.67e-16 264-291 IPB000907B 14.10 2.50e-13 160-175 PR00467A 8.38 3.29e-13 11-28 PR00467B 14.98 5.88e-12 57-76 IPB000907A 16.20 4.21e-10 94-103 PR00467C 9.34 7.65e-10 134-148
927	IPB001774	Delta serrate ligand	IPB001774C 18.25 1.71e-31 37-79 IPB001774D 19.23 3.32e-25 83-129
927	PR00011	Type III EGF-like signature IV	PR00011D 12.12 4.57e-12 39-57
927	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 1.00e-10 189-204 IPB001774C 18.25 2.15e-10 68-110
927	PR00010	Type II EGF-like signature III	PR00010C 6.98 3.90e-10 113-123
927	IPB000359	Cystine-knot domain	IPB000359A 23.24 4.86e-10 160-184
927	IPB000034	Laminin B	IPB000034C 12.97 6.42e-10 236-254 PR00011B 13.08 7.88e-10 39-57
927	IPB000561	EGF-like domain	IPB000561 4.89 9.25e-10 46-54
927	IPB001886	Laminin N-terminal (Domain VI)	IPB001886E 10.90 9.67e-10 44-60 PR00010A 12.91 1.27e-09 174-185 PR00010C 6.98 2.54e-09 194-204
927	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862F 29.39 2.65e-09 201-248 IPB000152 8.86 6.21e-09 108-123 PR00011A 14.05 6.88e-09 39-57
927	PR01217	Proline rich extensin signature VII	PR01217G 4.02 7.79e-09 252-277 IPB001862F 29.39 8.53e-09 163-210 IPB000034A 22.21 9.00e-09 96-131 IPB000152 8.86 9.29e-09 227-242
927	IPB001762	Disintegrin	IPB001762A 23.93 9.65e-09 126-166
928	PR00456	Ribosomal protein P2 signature V	PR00456E 3.08 7.80e-09 1-15
930	IPB001248	"Permeases for cytosine/purines, uracil, thiamine, allantoin"	IPB001248A 28.27 5.94e-10 238-273
930	IPB000390	"Integral membrane protein, DUF7"	IPB000390B 26.91 6.96e-10 217-271
931	IPB001359	Synapsin	IPB001359H 22.58 9.63e-10 47-97
932	PR00336	Lysosome-associated membrane glycoprotein signature IV	PR00336D 10.26 5.99e-09 2-24
933	IPB002467	"Methionine aminopeptidase, subfamily 1"	IPB002467C 17.56 2.29e-30 169-197 IPB002467B 12.68 2.50e-23 143-164 IPB002467F 18.38 1.71e-21 299-329
933	PR00599	Methionine aminopeptidase-1 signature II	PR00599B 10.21 8.00e-17 173-189 IPB002467D 14.78 5.50e-15 242-267 PR00599A 11.84 9.63e-14 151-164 IPB002467E 11.05 7.75e-12 275-287 PR00599D 14.43 5.03e-10 273-285 IPB002467A 15.75 2.87e-09 115-132
933	IPB001131	Proline dipeptidase	IPB001131D 11.56 5.18e-09 275-288 IPB001131B 18.96 8.10e-09 173-194
934	IPB001463	Sodium:alanine symporter family	IPB001463A 16.70 5.87e-09 174-224
938	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478A 11.55 5.09e-09 119-129 IPB001478B 6.12 1.00e-08 137-146
940	PR01286	Orphan nuclear receptor NOR1 signature V	PR01286E 5.27 9.26e-09 307-328
941	IPB000998	MAM domain	IPB000998D 18.66 1.96e-15 527-550
941	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.77e-15 237-256
941	IPB000152	Aspartic acid and asparagine	IPB000152 8.86 2.89e-14 110-125

391

TABLE 3B

		hydroxylation site	
941	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 5.00e-14 192-203 IPB000152 8.86 1.00e-13 237-252 IPB000152 8.86 1.82e-13 192-207 IPB001881B 12.28 4.75e-13 110-121
941	IPB001774	Delta serrate ligand	IPB001774C 18.25 9.13e-13 72-114 IPB000998B 17.20 1.00e-12 410-422
941	PR00020	MAM domain signature I	PR00020A 20.48 2.88e-11 408-426 IPB000998C 18.63 5.30e-11 464-479 IPB001881B 12.28 8.58e-11 237-248
941	PR00907	Thrombomodulin signature II	PR00907B 11.50 2.44e-10 144-160
941	IPB000561	EGF-like domain	IPB000561 4.89 3.25e-10 81-89
941	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 5.35e-10 242-252 IPB000033B 7.05 5.97e-09 197-207
941	IPB000167	Dehydrin	IPB000167A 8.58 7.14e-09 324-351
941	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 9.79e-09 159-179
942	IPB000998	MAM domain	IPB000998D 18.66 1.96e-15 532-555
942	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.77e-15 242-261
942	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 2.89e-14 115-130
942	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 5.00e-14 197-208 IPB000152 8.86 1.00e-13 242-257 IPB000152 8.86 1.82e-13 197-212 IPB001881B 12.28 4.75e-13 115-126
942	IPB001774	Delta serrate ligand	IPB001774C 18.25 9.13e-13 77-119 IPB000998B 17.20 1.00e-12 415-427
942	PR00020	MAM domain signature I	PR00020A 20.48 2.88e-11 413-431 IPB000998C 18.63 5.30e-11 469-484 IPB001881B 12.28 8.58e-11 242-253
942	PR00907	Thrombomodulin signature II	PR00907B 11.50 2.44e-10 149-165
942	IPB000561	EGF-like domain	IPB000561 4.89 3.25e-10 86-94
942	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 5.35e-10 247-257
942	PR01256	Otx1 transcription factor signature II	PR01256B 5.92 2.01e-09 23-35 IPB000033B 7.05 5.97e-09 202-212 PR01256B 5.92 6.46e-09 24-36
942	IPB000167	Dehydrin	IPB000167A 8.58 7.14e-09 329-356
942	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 9.79e-09 164-184
943	IPB002893	MYND zinc finger (ZnF) domain	IPB002893 16.28 4.52e-17 986-1004
943	IPB000313	PWWP domain	IPB000313A 8.15 6.88e-15 276-290
943	IPB001487	Bromodomain	IPB001487B 17.44 1.32e-13 202-223 IPB001487A 11.44 9.33e-12 178-196
943	IPB002219	Phorbol esters/diacylglycerol binding domain	IPB002219B 12.53 5.14e-10 94-109
943	PR00503	Bromodomain signature II	PR00503B 10.44 7.38e-09 177-193
943	IPB002889	WSC domain	IPB002889C 9.89 8.12e-09 762-783 IPB002889B 11.76 9.91e-09 744-790
944	IPB000313	PWWP domain	IPB000313A 8.15 6.88e-15 276-290
944	IPB001487	Bromodomain	IPB001487B 17.44 1.32e-13 202-223 IPB001487A 11.44 9.33e-12 178-196
944	IPB002219	Phorbol esters/diacylglycerol binding domain	IPB002219B 12.53 5.14e-10 94-109
944	PR00503	Bromodomain signature II	PR00503B 10.44 7.38e-09 177-193
945	IPB002893	MYND zinc finger (ZnF) domain	IPB002893 16.28 4.52e-17 1032-1050
945	IPB000313	PWWP domain	IPB000313A 8.15 6.88e-15 276-290
945	IPB001487	Bromodomain	IPB001487B 17.44 1.32e-13 202-223 IPB001487A 11.44 9.33e-12 178-196

392

TABLE 3B

945	IPB002219	Phorbol esters/diacylglycerol binding domain	IPB002219B 12.53 5.14e-10 94-109
945	PR00503	Bromodomain signature II	PR00503B 10.44 7.38e-09 177-193
945	IPB002889	WSC domain	IPB002889C 9.89 8.12e-09 762-783 IPB002889B 11.76 9.91e-09 744-790
946	IPB002893	MYND zinc finger (ZnF) domain	IPB002893 16.28 4.52e-17 1037-1055
946	IPB000313	PWWP domain	IPB000313A 8.15 6.88e-15 281-295
946	IPB001487	Bromodomain	IPB001487B 17.44 1.32e-13 207-228 IPB001487A 11.44 9.33e-12 183-201
946	IPB002219	Phorbol esters/diacylglycerol binding domain	IPB002219B 12.53 5.14e-10 99-114
946	PR00503	Bromodomain signature II	PR00503B 10.44 7.38e-09 182-198
946	IPB002889	WSC domain	IPB002889C 9.89 8.12e-09 767-788 IPB002889B 11.76 9.91e-09 749-795
950	PR00169	Potassium channel signature VII	PR00169G 11.30 5.96e-11 467-489
950	PR01333	Two pore domain K ⁺ channel signature I	PR01333A 18.74 7.08e-10 479-507 PR01333B 10.39 5.95e-09 482-491
950	PR00206	Connexin signature VI	PR00206F 15.67 6.01e-09 498-521
951	IPB001762	Disintegrin	IPB001762A 23.93 4.33e-23 441-481
951	IPB002870	Reprolysin family propeptide	IPB002870B 24.73 3.54e-20 114-152
951	PR00289	Disintegrin signature I	PR00289A 14.29 1.16e-14 457-476 IPB002870F 18.81 3.03e-14 385-409 IPB002870E 11.90 2.46e-12 344-356 IPB001762B 10.06 3.40e-12 488-498 IPB001762A 23.93 9.20e-11 409-449
951	IPB000130	"Neutral zinc metallopeptidases, zinc-binding region"	IPB000130 5.86 1.56e-10 342-352
951	PR00138	Matrixin signature IV	PR00138D 14.57 2.54e-10 342-367 IPB002870D 16.31 4.77e-10 310-325
951	PR00480	Astacin family signature II	PR00480B 14.35 5.57e-10 337-355
951	PR00436	Interleukin-8 signature I	PR00436A 15.20 7.43e-10 5-28
951	IPB001818	Matrixin	IPB001818D 14.91 1.72e-09 336-367 PR00289B 11.74 3.80e-09 486-498 IPB002870A 12.22 6.54e-09 68-84
951	PR01236	Tumour necrosis factor beta (lymphotoxin-alpha) signature I	PR01236A 4.92 7.49e-09 17-33 IPB002870C 11.01 9.64e-09 278-288
953	IPB000906	ZU5 domain	IPB000906E 22.11 5.55e-11 248-288
953	PR01415	Ankyrin repeat signature I	PR01415A 12.73 6.46e-11 251-263 IPB000906D 23.89 6.59e-11 316-370 PR01415A 12.73 7.11e-11 184-196 PR01415A 12.73 7.43e-11 152-164 IPB000906F 35.93 5.85e-10 194-247 PR01415B 10.23 5.88e-09 263-275 IPB000906G 25.85 6.69e-09 330-378
953	PR00898	Vasopressin V2 receptor signature II	PR00898B 4.91 7.69e-09 46-60 IPB000906A 22.49 7.84e-09 177-219
954	IPB000471	"Interferon alpha, beta and delta family"	IPB000471A 27.36 3.61e-32 45-98
954	PR00266	Interferon alpha and beta subunit signature I	PR00266A 13.41 9.59e-14 67-79
955	PR01136	Gap junction alpha-6 protein (Cx45) signature I	PR01136A 6.68 5.05e-09 203-209
956	PR00081	Glucose/ribitol dehydrogenase family signature VI	PR00081F 13.94 5.50e-13 152-172 PR00081A 10.07 5.67e-13 34-51 PR00081B 8.91 5.66e-11 108-119
956	PR01397	"2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase signature VI"	PR01397F 12.91 9.53e-11 168-187

393

TABLE 3B

956	PR00080	Short-chain dehydrogenase/reductase (SDR) superfamily signature I	PR00080A 7.98 3.73e-09 108-119 PR01397A 13.33 4.65e-09 39-56
958	IPB000560	Histidine acid phosphatase	IPB000560 17.02 7.55e-13 30-52
958	PR00885	Bacterial general secretion pathway protein H signature II	PR00885B 8.16 9.14e-10 394-408
958	PR01319	Glial cell line-derived neurotrophic factor receptor alpha 3 signature I	PR01319A 3.85 3.93e-09 10-22
959	IPB000215	Serpins	IPB000215D 15.35 7.00e-22 224-250 IPB000215E 15.36 6.06e-18 305-329 IPB000215C 13.90 4.75e-17 122-136 IPB000215B 9.87 3.84e-12 95-107
960	IPB000215	Serpins	IPB000215D 15.35 7.00e-22 292-318 IPB000215A 13.01 4.18e-20 73-96 IPB000215E 15.36 6.06e-18 373-397 IPB000215C 13.90 5.82e-11 190-204
961	IPB000215	Serpins	IPB000215D 15.35 7.00e-22 292-318 IPB000215A 13.01 4.18e-20 73-96 IPB000215E 15.36 6.06e-18 373-397 IPB000215C 13.90 4.75e-17 190-204 IPB000215B 9.87 3.84e-12 163-175
962	IPB000215	Serpins	IPB000215A 13.01 4.18e-20 73-96 IPB000215E 15.36 6.06e-18 373-397 IPB000215C 13.90 4.75e-17 208-222 IPB000215B 9.87 3.84e-12 181-193
964	IPB001762	Disintegrin	IPB001762A 23.93 4.33e-23 457-497
964	IPB002870	Reprolysin family propeptide	IPB002870F 18.81 2.35e-19 402-426 IPB002870E 11.90 3.37e-16 366-378 IPB002870B 24.73 8.16e-16 145-183
964	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 8.05e-14 789-813
964	PR00289	Disintegrin signature I	PR00289A 14.29 2.80e-13 473-492 IPB000135D 2.13 6.08e-13 788-812 IPB000135D 2.13 9.08e-13 785-809 IPB000135D 2.13 2.30e-12 786-810 IPB000135D 2.13 6.10e-12 787-811 IPB000135D 2.13 6.75e-12 790-814
964	IPB001580	Calreticulin family	IPB001580F 2.93 5.50e-11 794-803 IPB002870A 12.22 8.80e-11 100-116 IPB000135D 2.13 3.64e-10 783-807 IPB001762B 10.06 4.86e-10 504-514 IPB001580F 2.93 4.94e-10 801-810 IPB001580F 2.93 4.94e-10 802-811 IPB000135D 2.13 6.09e-10 784-808 IPB000135D 2.13 9.71e-10 782-806 IPB002870D 16.31 9.71e-10 332-347 IPB001580F 2.93 1.00e-09 798-807
964	IPB000130	"Neutral zinc metallopeptidases, zinc-binding region"	IPB000130 5.86 1.86e-09 364-374 PR00289B 11.74 1.89e-09 502-514 IPB002870C 11.01 3.16e-09 300-310
964	IPB003191	Guanylate-binding protein	IPB003191N 9.33 3.37e-09 779-809
964	PR00480	Astacin family signature II	PR00480B 14.35 3.45e-09 359-377
964	IPB001422	Neuromodulin (GAP-43)	IPB001422C 16.82 4.49e-09 777-812
965	IPB000329	Uteroglobin family	IPB000329A 11.99 3.57e-10 1-16
965	PR00486	Uteroglobin signature I	PR00486A 6.53 9.03e-09 2-16
966	IPB000407	GDA1/CD39 family of nucleoside phosphatase	IPB000407C 15.11 5.50e-24 175-197 IPB000407D 11.44 2.16e-14 216-229 IPB000407B 8.75 3.86e-13 132-143

TABLE 3B

			IPB000407F 16.53 3.89e-12 422-436 IPB000407A 11.93 5.30e-12 56-67 IPB000407E 19.08 8.20e-11 342-358 IPB000407G 17.95 8.20e-11 455-469
967	IPB001073	Complement C1q protein	IPB001073B 20.88 5.78e-23 96-130 IPB001073C 13.07 4.50e-13 163-182 IPB001073A 22.14 6.55e-13 42-76
967	PR00007	Complement C1Q domain signature II	PR00007B 15.63 9.56e-13 116-135 IPB001073D 7.60 1.00e-11 195-204 PR00007D 9.66 2.00e-11 193-203 PR00007C 16.13 7.38e-11 163-184 PR00007A 20.64 9.32e-10 89-115
970	IPB000721	Gag gene protein p24 (core nucleocapsid protein)	IPB000721E 14.33 1.57e-12 525-538
970	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.09e-11 206-243 IPB003006A 17.51 1.00e-10 160-182
970	IPB001020	Histidine phosphorylation site in HPr protein	IPB001020B 19.38 4.53e-09 378-416
971	IPB001759	Pentaxin family	IPB001759D 18.25 4.67e-33 409-447
971	PR00895	Pentaxin signature V	PR00895E 12.84 4.19e-18 417-436 PR00895D 14.46 2.38e-17 397-416 PR00895C 12.82 3.18e-17 370-388 IPB001759C 13.49 4.30e-17 370-388 IPB001759A 29.51 1.82e-14 113-147 PR00895A 14.28 8.83e-13 305-319 PR00895B 14.42 1.45e-12 327-341 IPB001759B 14.85 3.30e-11 327-341 IPB001759E 18.14 5.34e-11 459-473 PR00895F 15.89 9.50e-11 436-450
973	IPB002889	WSC domain	IPB002889B 11.76 5.15e-13 453-499 IPB002889B 11.76 1.55e-12 445-491 IPB002889B 11.76 4.18e-12 458-504
973	IPB001871	bZIP (Basic-leucine zipper) transcription factor family	IPB001871 8.42 8.65e-12 633-645 IPB002889B 11.76 8.79e-12 447-493 IPB002889B 11.76 9.89e-12 440-486 IPB002889B 11.76 2.59e-11 439-485 IPB002889B 11.76 4.49e-11 441-487 IPB002889B 11.76 5.13e-11 454-500 IPB002889B 11.76 5.87e-11 437-483 IPB002889B 11.76 6.72e-11 448-494
973	PR00043	Jun transcription factor signature II	PR00043B 8.71 8.92e-11 633-649
973	PR01449	Calcium-activated BK potassium channel alpha subunit signature VIII	PR01449H 2.34 9.85e-11 468-483 IPB002889B 11.76 2.19e-10 449-495 IPB002889B 11.76 2.58e-10 443-489 IPB002889B 11.76 3.87e-10 456-502 IPB002889B 11.76 4.46e-10 452-498 IPB002889B 11.76 6.44e-10 444-490
973	IPB002546	Myogenic Basic domain	IPB002546E 13.48 9.04e-10 464-481 IPB002889B 11.76 9.41e-10 457-503 IPB002889B 11.76 1.00e-09 461-507 IPB002889B 11.76 1.28e-09 436-482
973	IPB000684	Eukaryotic RNA polymerase II heptapeptide repeat	IPB000684L 3.49 2.10e-09 445-487 IPB002889C 9.89 2.21e-09 466-487 PR01449H 2.34 2.50e-09 469-484 PR01449H 2.34 2.50e-09 472-487 PR01449H 2.34 2.59e-09 466-481 PR01449H 2.34 2.59e-09 467-482 PR01449H 2.34 3.03e-09 463-478

TABLE 3B

			IPB002889B 11.76 4.09e-09 438-484 PR01449H 2.34 4.18e-09 461-476 PR01449H 2.34 4.18e-09 464-479 PR01449H 2.34 4.35e-09 473-488 IPB002889B 11.76 4.47e-09 455-501 PR01449H 2.34 4.53e-09 453-468 IPB002889B 11.76 5.13e-09 442-488 IPB002889B 11.76 5.31e-09 431-477 IPB002546E 13.48 5.50e-09 469-486 IPB002889B 11.76 6.62e-09 463-509 IPB002889B 11.76 7.19e-09 462-508 IPB002889B 11.76 8.69e-09 450-496 IPB000684L 3.49 8.83e-09 447-489
977	IPB001359	Synapsin	IPB001359H 22.58 1.95e-15 545-595
977	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 3.00e-13 2087-2112 IPB000822 14.67 1.86e-11 2476-2501 IPB001359H 22.58 4.46e-11 539-589 IPB000822 14.67 5.29e-11 2362-2387 IPB000822 14.67 6.57e-11 472-497 IPB000822 14.67 8.71e-11 2253-2278
977	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 9.02e-11 540-554 PR00049D 0.00 9.17e-11 541-555
977	IPB003861	E4 protein	IPB003861B 9.06 1.43e-10 547-561
977	IPB002999	Tudor domain	IPB002999C 10.33 2.00e-10 546-555 IPB000822 14.67 2.29e-10 110-135 IPB001359H 22.58 2.67e-10 537-587
977	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 3.45e-10 2473-2486 IPB001359H 22.58 5.08e-10 551-601 IPB001359H 22.58 5.36e-10 541-591
977	PR01217	Proline rich extensin signature VII	PR01217G 4.02 6.94e-10 545-570 IPB000822 14.67 1.00e-09 602-627 PR00049D 0.00 2.98e-09 538-552 PR00049D 0.00 3.29e-09 539-553
977	IPB002000	Lysosome-associated membrane glycoprotein (Lamp)	IPB002000D 5.87 3.72e-09 192-205 PR00049D 0.00 3.90e-09 543-557
977	IPB000413	Integrins alpha chain	IPB000413A 13.51 4.33e-09 1509-1519 IPB001359H 22.58 4.41e-09 547-597 PR00049D 0.00 4.81e-09 537-551
977	PR00021	Small proline-rich protein signature I	PR00021A 3.31 5.38e-09 538-550 IPB000822 14.67 5.50e-09 1894-1919 IPB000822 14.67 5.88e-09 1579-1604 IPB001359H 22.58 5.89e-09 543-593 PR00049D 0.00 6.03e-09 191-205 IPB000822 14.67 6.62e-09 1662-1687
977	PR00239	Molluscan rhodopsin C-terminal tail signature V	PR00239E 1.29 6.97e-09 542-553 IPB000822 14.67 7.00e-09 2053-2078 IPB001359H 22.58 7.03e-09 546-596 PR00048B 5.52 7.50e-09 2100-2109 IPB002999B 7.50 7.55e-09 545-553 IPB002999B 7.50 7.55e-09 546-554 IPB000822 14.67 8.12e-09 2116-2141 IPB000822 14.67 8.50e-09 1267-1292
977	PR00776	Hemoglobinase (C13) cysteine protease signature IV	PR00776D 11.72 8.62e-09 2447-2466 IPB001359H 22.58 8.95e-09 558-608 IPB002000D 5.87 9.49e-09 542-555
977	PR00211	Glutelin signature II	PR00211B 0.86 9.92e-09 551-571 IPB000822 14.67 1.00e-08 1032-1057

396

TABLE 3B

980	PR00834	HtrA/DegQ protease family signature III	PR00834C 15.48 6.81e-20 237-261 PR00834D 11.75 9.45e-18 275-292
980	IPB002350	Kazal-type serine protease inhibitor family	IPB002350 31.78 6.52e-17 73-113 PR00834B 10.17 6.63e-14 196-216 PR00834E 13.43 9.13e-13 297-314
980	IPB000867	Insulin-like growth factor-binding protein	IPB000867B 11.44 1.94e-12 23-39
980	IPB000126	"Serine proteases, V8 family"	IPB000126B 12.50 3.32e-12 280-296 PR00834F 11.11 3.25e-11 389-401 PR00834A 8.79 5.83e-11 175-187 IPB000126A 11.75 5.69e-10 173-188
980	PR00290	Kazal-type serine protease inhibitor signature II	PR00290B 16.63 2.80e-09 84-95
980	PR00722	Chymotrypsin serine protease family (S1) signature III	PR00722C 10.74 4.10e-09 283-295
980	PR01424	Transforming growth factor beta 1 precursor signature I	PR01424A 6.58 8.24e-09 8-27
980	IPB001489	Heat-stable enterotoxin	IPB001489 13.51 8.78e-09 26-38
981	PR00792	Pepsin (A1) aspartic protease family signature I	PR00792A 11.02 5.32e-17 80-100
981	IPB001969	Eukaryotic and viral aspartic protease active site	IPB001969A 16.37 5.15e-13 87-103 PR00792D 11.77 1.00e-12 395-410 PR00792C 8.65 6.29e-12 312-323 IPB001969A 16.37 7.00e-10 310-326
982	IPB000917	Sulfatase	IPB000917A 9.52 5.26e-10 44-55
984	IPB000834	"Zinc carboxypeptidases, carboxypeptidase A metalloprotease (M14) family"	IPB000834B 13.51 2.50e-17 103-117
984	PR00765	Carboxypeptidase A metalloprotease (M14) family signature II	PR00765B 14.48 1.39e-15 99-113 IPB000834C 17.20 2.80e-15 172-188 IPB000834G 14.46 4.50e-15 318-333 IPB000834D 18.95 4.72e-12 199-225 PR00765D 14.06 9.45e-12 233-246 PR00765C 10.88 1.82e-10 179-187 IPB000834F 12.40 4.21e-10 285-297 IPB000834E 9.80 2.15e-09 228-242
985	IPB000834	"Zinc carboxypeptidases, carboxypeptidase A metalloprotease (M14) family"	IPB000834B 13.51 2.50e-17 103-117
985	PR00765	Carboxypeptidase A metalloprotease (M14) family signature II	PR00765B 14.48 1.39e-15 99-113 IPB000834C 17.20 2.80e-15 172-188 IPB000834G 14.46 4.50e-15 318-333 IPB000834D 18.95 4.72e-12 199-225 PR00765D 14.06 9.45e-12 233-246 PR00765C 10.88 1.82e-10 179-187 IPB000834F 12.40 4.21e-10 285-297 IPB000834E 9.80 2.15e-09 228-242
986	IPB002871	NifU-like N terminal domain	IPB002871C 16.51 1.60e-33 81-113 IPB002871D 14.11 6.87e-21 131-153 IPB002871A 14.39 2.17e-17 35-50 IPB002871B 12.43 6.79e-14 62-74
990	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 8.29e-11 94-119
990	PR00048	C2H2-type zinc finger signature II	PR00048B 5.52 9.50e-09 107-116
991	IPB003527	MAP kinase	IPB003527D 21.53 5.58e-23 185-226 IPB003527G 17.26 8.24e-22 285-322 IPB003527C 14.70 3.05e-19 124-172
991	IPB001245	Tyrosine kinase catalytic domain	IPB001245A 22.45 5.50e-17 132-172

397

TABLE 3B

991	IPB000959	POLO box duplicated region	IPB000959B 15.68 7.19e-17 116-156 IPB001245B 21.68 1.39e-15 192-230
991	IPB001772	Kinase associated domain 1	IPB001772C 20.66 3.92e-14 127-157
991	IPB000095	PAK-box /P21-Rho-binding	IPB000095C 13.36 7.91e-13 46-82 IPB003527A 17.00 6.14e-12 26-51
991	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861G 13.73 7.44e-12 194-243
991	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 5.91e-11 188-229 IPB003527B 11.51 9.15e-11 98-116
991	PR00109	Tyrosine kinase catalytic domain signature II	PR00109B 11.07 9.10e-10 139-157 IPB000961C 15.48 8.83e-09 139-173
992	PR01432	Rabaptin signature XI	PR01432K 2.19 8.43e-09 976-998
994	IPB001073	Complement C1q protein	IPB001073B 20.88 7.26e-29 175-209
994	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 8.93e-27 75-127
994	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 2.83e-26 74-127 IPB000885B 19.15 7.37e-23 80-133 IPB001442A 26.12 7.39e-23 72-124 IPB000885B 19.15 8.75e-23 77-130 IPB000885A 11.46 1.79e-21 82-119 IPB001073A 22.14 2.24e-21 78-112 IPB000885A 11.46 3.84e-21 79-116 IPB000885A 11.46 5.11e-21 76-113 IPB000885B 19.15 5.89e-21 71-124 IPB000885B 19.15 7.56e-21 68-121 IPB001442A 26.12 8.15e-21 66-118 IPB001442A 26.12 8.40e-21 69-121 IPB000885B 19.15 2.97e-20 62-115 IPB001442A 26.12 3.72e-20 78-130 IPB000885A 11.46 4.00e-20 70-107 IPB001442A 26.12 5.62e-20 63-115
994	PR00007	Complement C1Q domain signature I	PR00007A 20.64 6.54e-20 168-194 IPB000885A 11.46 8.20e-20 73-110 IPB001442A 26.12 9.64e-20 84-136 IPB001442A 26.12 3.69e-19 87-139 IPB001442A 26.12 5.09e-19 60-112 IPB001442A 26.12 7.43e-19 81-133 IPB000885B 19.15 3.81e-18 83
994	IPB000817	Prion protein	IPB000817A 8.34 9.51e-10 76-118 IPB001442B 12.38 1.00e-09 106-126 IPB000885A 11.46 4.12e-09 58-95 IPB001442B 12.38 5.01e-09 97-117 IPB000817A 8.34 6.12e-09 77-119 IPB001442B 12.38 7.32e-09 73-93 IPB000885A 11.46 7.34e-09 106-143 IPB001442B 12.38 7.93e-09 70-90 IPB000885A 11.46 8.16e-09 55-92 IPB000885B 19.15 8.77e-09 101-154 IPB000817A 8.34 9.43e-09 65-107
996	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.24e-10 311-348
997	IPB001895	Guanine-nucleotide dissociation stimulators CDC25 family	IPB001895C 20.83 7.84e-30 1077-1112 IPB001895D 18.68 1.00e-20 1174-1197
997	IPB001331	Guanine-nucleotide dissociation stimulators CDC24 family	IPB001331C 16.09 1.00e-18 377-402 IPB001895B 16.80 3.10e-15 1005-1025 IPB001331B 19.33 7.00e-09 326-341
999	IPB002360	Involucrin	IPB002360C 15.36 3.70e-09 198-239

398

TABLE 3B

999	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 3.91e-09 202-226
999	PR00169	Potassium channel signature I	PR00169A 17.48 5.50e-09 68-87
999	PR01083	Lymphocyte-specific protein signature I	PR01083A 8.60 9.61e-09 214-237
1001	IPB000492	Protamine 2 (PRM2)	IPB000492B 5.26 5.11e-09 788-822
1001	IPB000221	Protamine P1	IPB000221 5.48 7.46e-09 945-971 IPB000221 5.48 8.85e-09 831-857
1002	IPB003403	Herpesvirus immediate early protein	IPB003403E 17.25 6.47e-10 52-79
1002	IPB001841	RING finger	IPB001841 10.69 3.84e-09 126-135
1002	IPB000492	Protamine 2 (PRM2)	IPB000492B 5.26 5.11e-09 997-1031
1002	IPB000221	Protamine P1	IPB000221 5.48 7.46e-09 1154-1180 IPB000221 5.48 8.85e-09 1040-1066
1003	PR00320	G protein beta WD-40 repeat signature I	PR00320A 13.15 4.32e-12 1132-1146 PR00320C 12.32 3.14e-11 1132-1146 PR00320B 12.82 7.55e-11 1132-1146 PR00320A 13.15 8.92e-10 1091-1105 PR00320C 12.32 1.33e-09 1091-1105
1003	IPB001680	G-protein beta WD-40 repeats	IPB001680 10.43 1.45e-09 1134-1145 PR00320B 12.82 2.24e-09 1091-1105 PR00320A 13.15 4.86e-09 789-803
1003	PR01472	Intercellular adhesion molecule/vascular cell adhesion molecule-1 signature I	PR01472A 16.78 9.82e-09 1154-1170
1004	IPB000433	ZZ Zinc finger	IPB000433 14.10 8.20e-18 21-37
1004	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.86e-10 80-105
1006	IPB000008	C2 domain	IPB000008C 23.37 8.91e-26 323-362 IPB000008D 14.83 1.23e-12 378-396 IPB000008B 17.91 3.09e-09 281-298 IPB000008E 14.84 3.90e-09 401-411
1007	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 5.91e-11 877-901 IPB000135D 2.13 7.44e-11 885-909 IPB000135D 2.13 7.85e-11 887-911 IPB000135D 2.13 3.05e-10 883-907 IPB000135D 2.13 5.11e-10 881-905 IPB000135D 2.13 8.14e-10 888-912 IPB000135D 2.13 2.27e-09 876-900 IPB000135D 2.13 2.27e-09 882-906 IPB000135D 2.13 2.36e-09 880-904
1007	PR00806	Vinculin signature IV	PR00806D 11.95 3.78e-09 564-579 IPB000135D 2.13 3.91e-09 874-898 IPB000135D 2.13 4.45e-09 889-913 IPB000135D 2.13 6.36e-09 884-908 IPB000135D 2.13 7.00e-09 879-903 IPB000135D 2.13 7.18e-09 886-910 IPB000135D 2.13 9.27e-09 920-944
1008	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 8.85e-21 560-584 IPB000135D 2.13 2.47e-19 559-583 IPB000135D 2.13 7.87e-19 561-585 IPB000135D 2.13 8.53e-19 563-587 IPB000135D 2.13 9.35e-19 558-582 IPB000135D 2.13 7.25e-18 564-588 IPB000135D 2.13 7.43e-17 55
1008	IPB003403	Herpesvirus immediate early protein	IPB003403E 17.25 6.81e-10 560-587
1008	IPB003874	CDC45-like protein	IPB003874C 5.49 1.24e-09 571-582
1008	IPB001990	Granins (chromogranin or secretogranin)	IPB001990C 33.59 3.49e-09 538-585

399

TABLE 3B

1008	IPB000637	HMG-I and HMG-Y DNA-binding domain (A+T-hook)	IPB000637B 14.21 5.64e-09 568-586 IPB000135D 2.13 6.09e-09 545-569
1008	IPB001580	Calreticulin family	IPB001580F 2.93 9.10e-09 573-582
1009	PR00405	HIV Rev interacting protein signature II	PR00405B 10.10 2.93e-17 281-298 PR00405A 18.83 3.86e-14 262-281
1009	PR00452	SH3 domain signature II	PR00452B 11.47 9.70e-10 895-910 PR00405C 18.05 3.95e-09 302-323
1009	IPB003134	Repeat in HS1/Cortactin	IPB003134H 12.06 4.27e-09 880-929
1009	PR00910	Luteovirus ORF6 protein signature I	PR00910A 2.74 8.71e-09 335-347
1011	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 218-255 IPB003006B 20.23 9.14e-12 511-548 IPB003006B 20.23 1.00e-11 318-355
1011	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.23e-11 617-640 IPB003006B 20.23 6.40e-10 124-161 IPB003006B 20.23 9.64e-10 610-647 IPB003006B 20.23 7.58e-09 21-58 IPB003006B 20.23 8.62e-09 416-453 PR01536C 19.92 9.19e-09 225-248
1015	IPB002048	EF-hand family	IPB002048 7.91 2.29e-11 147-159
1015	PR00450	Recoverin family signature III	PR00450C 11.99 1.58e-09 33-54 IPB002048 7.91 8.58e-09 74-86
1016	IPB003846	Uncharacterized protein family UPF0061	IPB003846E 18.41 1.00e-40 136-174 IPB003846F 24.67 9.36e-31 175-210 IPB003846D 28.31 1.61e-17 52-94 IPB003846G 13.31 5.09e-09 268-278
1017	IPB003846	Uncharacterized protein family UPF0061	IPB003846C 15.01 1.00e-40 176-219 IPB003846E 18.41 1.00e-40 468-506 IPB003846F 24.67 9.36e-31 507-542 IPB003846D 28.31 7.86e-25 235-277 IPB003846B 13.03 2.00e-11 148-159 IPB003846A 5.99 3.25e-11 140-146 IPB003846G 13.31 5.09e-09 600-610
1017	PR01548	Meiotic recombination protein rec114 signature I	PR01548A 10.11 6.52e-09 238-258
1018	PR00237	Rhodopsin-like GPCR superfamily signature V	PR00237E 13.03 3.12e-16 236-259
1018	PR00238	Opsin signature II	PR00238B 16.77 4.52e-14 208-220 PR00237D 9.76 7.92e-14 186-207 PR00237B 12.45 1.39e-13 105-126 PR00237F 14.34 1.67e-13 294-318 PR00237C 14.77 2.00e-13 150-172 PR00237G 19.23 4.00e-13 332-358
1018	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276B 4.97 6.62e-13 244-255 PR00237A 9.81 7.00e-12 72-96 IPB000276A 11.56 5.24e-11 164-175 IPB000276D 9.40 4.52e-10 342-358 PR00238A 12.47 6.65e-09 93-105
1018	PR00667	Retinal pigment epithelium-retinal GPCR signature II	PR00667B 10.86 8.80e-09 91-106
1019	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 2.80e-13 378-391 PR00019B 11.42 2.33e-10 131-144 PR00019B 11.42 6.33e-10 375-388 PR00019B 11.42 3.73e-09 225-238 PR00019B 11.42 4.00e-09 249-262 PR00019A 11.72 4.55e-09 252-265 PR00019A 11.72 8.09e-09 134-147
1021	IPB001895	Guanine-nucleotide dissociation	IPB001895C 20.83 3.00e-28 984-1019

400

TABLE 3B

		stimulators CDC25 family	IPB001895D 18.68 8.56e-17 1082-1105 IPB001895B 16.80 4.30e-15 913-933
1021	IPB000595	Cyclic nucleotide-binding domain	IPB000595B 15.72 6.40e-11 355-378
1021	IPB003351	Dishevelled specific domain	IPB003351F 12.17 4.43e-10 615-641
1021	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478B 6.12 3.25e-09 625-634
1021	PR00834	HtrA/DegQ protease family signature VI	PR00834F 11.11 6.03e-09 621-633
1022	IPB001895	Guanine-nucleotide dissociation stimulators CDC25 family	IPB001895C 20.83 3.00e-28 934-969 IPB001895D 18.68 8.56e-17 1032-1055 IPB001895B 16.80 4.30e-15 863-883
1022	IPB000595	Cyclic nucleotide-binding domain	IPB000595B 15.72 6.40e-11 305-328
1022	IPB003351	Dishevelled specific domain	IPB003351F 12.17 4.43e-10 565-591
1022	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478B 6.12 3.25e-09 575-584
1022	PR00834	HtrA/DegQ protease family signature VI	PR00834F 11.11 6.03e-09 571-583
1024	PR00907	Thrombomodulin signature VIII	PR00907H 1.34 7.64e-09 376-400
1025	PR00907	Thrombomodulin signature VIII	PR00907H 1.34 7.64e-09 338-362
1027	IPB003452	Stem cell factor	IPB003452A 12.58 1.00e-40 1-41 IPB003452D 16.80 1.00e-40 173-211 IPB003452C 13.68 6.76e-37 131-164 IPB003452B 19.11 2.09e-18 53-101 IPB003452B 19.11 8.06e-17 43-91
1028	PR00205	Cadherin signature II	PR00205B 20.09 1.00e-19 150-179 PR00205D 12.22 9.31e-19 238-257 PR00205F 19.57 3.37e-17 316-342 PR00205B 20.09 6.67e-16 374-403 PR00205B 20.09 2.20e-15 259-288 PR00205A 17.38 6.82e-14 90-109 PR00205F 19.57 1.00e-13 97-123 PR00205F 19.57 6.70e-13 427-453
1028	IPB002126	Cadherin domain	IPB002126A 14.68 9.40e-13 101-117 IPB002126B 12.04 1.75e-12 247-264 PR00205G 13.05 4.30e-12 241-258 PR00205G 13.05 4.65e-11 499-516 IPB002126B 12.04 1.29e-10 138-155 PR00205E 10.82 2.17e-10 372-385 PR00205E 10.82 3.35e-10 257-270 IPB002126A 14.68 6.09e-10 431-447 PR00205D 12.22 6.55e-10 496-515 PR00205A 17.38 3.12e-09 420-439 PR00205D 12.22 5.33e-09 129-148
1029	PR00205	Cadherin signature II	PR00205B 20.09 1.00e-19 150-179 PR00205D 12.22 9.31e-19 238-257 PR00205F 19.57 3.37e-17 316-342 PR00205B 20.09 6.67e-16 374-403 PR00205B 20.09 2.20e-15 259-288 PR00205A 17.38 6.82e-14 90-109 PR00205F 19.57 1.00e-13 97-123 PR00205F 19.57 6.70e-13 427-453
1029	IPB002126	Cadherin domain	IPB002126A 14.68 9.40e-13 101-117 IPB002126B 12.04 1.75e-12 247-264 PR00205G 13.05 4.30e-12 241-258 PR00205G 13.05 4.65e-11 461-478 IPB002126B 12.04 1.29e-10 138-155 PR00205E 10.82 2.17e-10 372-385

401

TABLE 3B

			PR00205E 10.82 3.35e-10 257-270 IPB002126A 14.68 6.09e-10 431-447 PR00205D 12.22 6.55e-10 458-477 PR00205A 17.38 3.12e-09 420-439 PR00205D 12.22 5.33e-09 129-148
1030	PR00124	ATP synthase C subunit signature I	PR00124A 8.69 9.33e-10 41-60
1030	PR01131	Connexin36 (Cx36) signature II	PR01131B 3.45 3.17e-09 58-70 PR00124A 8.69 6.70e-09 43-62
1030	IPB003836	Glucokinase	IPB003836D 23.37 7.59e-09 48-81
1030	PR01516	Kv4.1 voltage-gated K ⁺ channel signature VII	PR01516G 4.80 8.98e-09 79-90
1031	IPB000180	Renal dipeptidase	IPB000180B 21.72 7.92e-34 242-281 IPB000180A 30.29 1.00e-33 172-215 IPB000180C 22.01 5.67e-27 287-321
1032	IPB002027	Amino acid permease	IPB002027D 22.00 4.13e-25 325-364 IPB002027C 19.67 2.74e-22 244-282 IPB002027A 18.88 3.77e-16 47-75 IPB002027B 12.67 7.97e-12 180-199
1033	IPB000559	Formate-tetrahydrofolate ligase	IPB000559C 13.05 1.00e-40 453-502 IPB000559F 12.78 1.00e-40 653-703 IPB000559G 15.54 1.00e-40 707-755 IPB000559D 22.27 4.33e-37 554-594 IPB000559E 17.08 7.39e-36 595-636 IPB000559K 15.77 8.96e-35 933-968 IPB000559B 12.60 2.88e-32 413-441 IPB000559J 17.25 5.94e-32 900-932 IPB000559H 20.31 2.72e-26 770-810 IPB000559A 24.17 6.11e-25 368-412 IPB000559I 15.05 6.35e-18 856-880
1033	PR00085	Tetrahydrofolate dehydrogenase/cyclohydrolase family signature III	PR00085C 13.81 5.70e-14 169-190 PR00085B 16.65 1.23e-09 136-163
1034	IPB000560	Histidine acid phosphatase	IPB000560 17.02 1.00e-11 378-400
1035	IPB001331	Guanine-nucleotide dissociation stimulators CDC24 family	IPB001331C 16.09 2.40e-12 911-936
1035	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 4.81e-09 1125-1139
1035	PR00834	HtrA/DegQ protease family signature VI	PR00834F 11.11 5.24e-09 82-94 PR00049D 0.00 5.73e-09 147-161
1035	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478B 6.12 7.19e-09 86-95
1035	IPB002532	Hantavirus glycoprotein G2	IPB002532J 16.97 8.37e-09 936-972
1035	PR00554	Adenosine A2B receptor signature II	PR00554B 12.52 8.85e-09 724-732
1037	PR00390	Phospholipase C signature I	PR00390A 14.24 6.34e-20 295-313
1037	IPB002048	EF-hand family	IPB002048 7.91 3.84e-09 147-159
1039	PR00245	Olfactory receptor signature III	PR00245C 14.65 5.26e-17 175-191 PR00245E 8.96 2.73e-13 282-293 PR00245B 13.73 1.39e-12 128-140 PR00245D 9.34 9.33e-11 235-244
1039	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 1.47e-10 117-128 PR00245A 10.98 8.80e-10 91-102 IPB000276D 9.40 9.61e-10 281-297
1039	PR00896	Vasopressin receptor signature II	PR00896B 9.36 5.50e-09 54-65
1039	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.70e-09 50-62
1039	PR00237	Rhodopsin-like GPCR superfamily signature II	PR00237B 12.45 7.16e-09 58-79 PR00237E 13.03 8.20e-09 198-221
1039	IPB003211	AmiS/UreI family transporter	IPB003211A 15.05 9.43e-09 27-66

402

TABLE 3B

1040	IPB003367	Thrombospondin type 3 repeat	IPB003367C 20.73 1.00e-40 428-478 IPB003367D 18.41 1.00e-40 479-521 IPB003367E 16.82 1.00e-40 522-569 IPB003367F 16.21 1.00e-40 580-629 IPB003367G 17.08 1.00e-40 630-671 IPB003367H 15.25 1.00e-40 672-704 IPB003367J 18.60 1.00
1040	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 4.79e-11 303-314 IPB003367E 16.82 5.67e-11 404-451 IPB003367C 20.73 5.96e-11 510-560 IPB003367E 16.82 6.83e-11 425-472 IPB003367C 20.73 2.38e-10 588-638 IPB003367C 20.73 6.35e-10 548-598
1040	IPB003129	Thrombospondin N-terminal -like domains	IPB003129B 23.30 7.86e-10 33-58 IPB003367C 20.73 8.46e-10 451-501 IPB003367E 16.82 8.88e-10 560-607 IPB003367C 20.73 6.20e-09 392-442 IPB003367E 16.82 6.95e-09 463-510
1040	IPB001774	Delta serrate ligand	IPB001774D 19.23 9.91e-09 226-272
1042	IPB000109	PTR peptide transporters (PTR2)	IPB000109D 25.09 6.67e-32 430-477 IPB000109B 29.23 4.18e-23 67-119 IPB000109A 10.85 3.79e-15 44-62 IPB000109C 8.21 7.00e-14 195-207
1042	PR00308	Type I antifreeze protein signature III	PR00308C 2.79 2.78e-09 20-29
1042	PR01471	Histamine H3 receptor signature II	PR01471B 12.38 9.63e-09 24-42
1043	IPB003104	Formin Homology 2 Domain	IPB003104B 18.83 6.87e-21 785-814 IPB003104C 20.33 1.27e-14 957-984
1043	IPB001073	Complement C1q protein	IPB001073A 22.14 3.25e-09 545-579
1043	IPB001359	Synapsin	IPB001359H 22.58 7.99e-09 553-603
1043	PR01471	Histamine H3 receptor signature V	PR01471E 5.41 8.14e-09 543-558
1044	IPB001909	KRAB box	IPB001909 17.37 6.32e-28 10-44
1044	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 9.10e-22 592-617 IPB000822 14.67 9.18e-21 228-253 IPB000822 14.67 5.50e-19 452-477 IPB000822 14.67 6.25e-19 284-309 IPB000822 14.67 7.23e-18 368-393 IPB000822 14.67 9.31e-18 144-169 IPB000822 14.67 2.29e-17 536-561 IPB000822 14.67 8.07e-17 480-505 IPB000822 14.67 9.36e-17 256-281 IPB000822 14.67 2.20e-16 340-365 IPB000822 14.67 5.20e-16 172-197 IPB000822 14.67 5.20e-16 200-225 IPB000822 14.67 5.80e-16 564-589 IPB000822 14.67 8.20e-16 396-421 IPB000822 14.67 8.80e-16 424-449 IPB000822 14.67 3.25e-15 508-533 IPB000822 14.67 4.94e-15 620-645
1044	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 5.50e-15 589-602 PR00048A 9.94 6.40e-15 253-266 IPB000822 14.67 1.00e-14 312-337 PR00048A 9.94 5.15e-14 533-546 PR00048A 9.94 6.79e-13 393-406 IPB000822 14.67 7.50e-13 116-141
1044	IPB001275	DM DNA binding domain	IPB001275 19.17 9.86e-13 580-619 PR00048A 9.94 1.53e-12 477-490 PR00048A 9.94 5.24e-12 561-574

403

TABLE 3B

			PR00048A 9.94 5.76e-12 225-238 IPB001275 19.17 8.66e-12 244-283 PR00048A 9.94 9.47e-12 281-294 PR00048A 9.94 1.00e-11 141-154
1044	IPB001222	TFIIS zinc ribbon domain	IPB001222 24.63 5.69e-09 116-152 PR00048B 5.52 7.00e-09 493-502 PR00048A 9.94 7.37e-09 421-434 PR00048A 9.94 9.25e-09 449-462 IPB001222 24.63 9.49e-09 144-180
1044	IPB002801	Aspartate carbamoyltransferase regulatory chain	IPB002801C 14.18 9.50e-09 254-270 PR00048B 5.52 9.50e-09 381-390
1046	IPB003137	Protease associated (PA) domain	IPB003137 22.40 2.50e-19 188-218
1048	IPB001627	Sema domain	IPB001627J 11.43 2.40e-11 403-419 IPB001627K 13.76 6.58e-11 477-489
1048	IPB002165	Plexin repeat	IPB002165D 14.72 7.91e-11 477-489
1049	IPB000243	Proteasome B-type subunit	IPB000243C 13.61 8.80e-09 52-62
1049	PR00766	Amiloride-sensitive amine oxidase signature VII	PR00766G 10.85 9.23e-09 91-111
1050	IPB001140	ABC transporter transmembrane region	IPB001140B 15.62 4.95e-14 138-176
1051	IPB000433	ZZ Zinc finger	IPB000433 14.10 8.20e-18 21-37
1051	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.86e-10 80-105
1052	IPB000353	"Class II histocompatibility antigen, beta chain, beta-1 domain"	IPB000353B 19.16 9.22e-14 133-182
1052	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.43e-12 86-123 IPB003006A 17.51 4.00e-11 154-176
1052	IPB001003	"MHC Class II, alpha chain, alpha-1 domain"	IPB001003B 14.72 5.40e-10 141-184
1053	PR00018	Kringle domain signature I	PR00018A 12.23 4.19e-09 36-51
1055	IPB001039	"Major histocompatibility complex protein, Class I"	IPB001039A 17.17 1.00e-40 15-68 IPB001039B 27.55 1.00e-40 96-147 IPB001039C 19.82 1.00e-40 177-230 IPB001039D 16.49 1.00e-40 255-309
1055	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.00e-30 261-298 IPB003006A 17.51 1.00e-21 224-246
1055	IPB000353	"Class II histocompatibility antigen, beta chain, beta-1 domain"	IPB000353B 19.16 7.65e-14 203-252
1055	IPB003363	Glycoprotein GG/GX	IPB003363E 13.35 8.75e-11 308-340
1055	IPB003705	Cobalt transport protein CbiN	IPB003705A 9.20 6.25e-09 316-332 IPB000353C 20.11 7.97e-09 254-308
1062	PR01382	Claudin-9 signature IV	PR01382D 12.38 1.11e-16 201-213
1062	IPB000729	PMP-22/EMP/MP20 family	IPB000729D 18.96 2.96e-16 160-187 IPB000729C 37.83 7.91e-16 80-132 PR01382A 12.00 1.17e-15 37-47
1062	PR01077	Claudin signature III	PR01077C 13.60 1.47e-14 63-73 PR01382C 5.67 5.14e-13 190-199 PR01382B 7.06 1.12e-12 91-100 PR01077B 14.12 1.00e-10 49-55 PR01077D 11.20 4.00e-10 146-152 PR01077A 9.72 8.16e-09 21-30
1064	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478B 6.12 5.50e-09 453-462 IPB001478B 6.12 7.75e-09 258-267
1066	IPB002659	Galactosyltransferase	IPB002659A 26.24 4.80e-11 92-133
1067	IPB001245	Tyrosine kinase catalytic domain	IPB001245A 22.45 7.60e-28 119-159
1067	IPB001772	Kinase associated domain 1	IPB001772C 20.66 9.25e-24 114-144
1067	IPB000961	Protein kinase C-terminal domain	IPB000961C 15.48 2.13e-22 126-160 IPB001772D 21.67 4.55e-17 186-225

TABLE 3B

1067	IPB000959	POLO box duplicated region	IPB000959B 15.68 8.60e-17 103-143
1067	IPB000095	PAK-box /P21-Rho-binding	IPB000095E 17.62 9.03e-17 127-172
1067	IPB003527	MAP kinase	IPB003527C 14.70 1.95e-16 111-159
1067	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861F 16.50 1.55e-15 120-174
1067	IPB000494	"Epidermal growth-factor receptor (EGFR), L domain"	IPB000494C 24.40 7.35e-14 113-159 IPB000959D 27.01 4.26e-13 226-278 IPB000961D 21.23 7.19e-13 175-216 IPB001245B 21.68 8.96e-13 179-217 IPB003527A 17.00 7.85e-11 18-43 IPB001772E 24.88 8.46e-11 233-272 IPB001772A 13.64 2.29e-10 9-40 IPB003527G 17.26 3.37e-09 245-282
1067	PR00109	Tyrosine kinase catalytic domain signature II	PR00109B 11.07 4.23e-09 126-144 IPB003527D 21.53 4.60e-09 172-213
1068	PR01254	Prostaglandin D synthase signature I	PR01254A 12.32 3.37e-29 31-54 PR01254D 13.80 7.97e-27 109-132 PR01254C 10.60 4.68e-22 74-92 PR01254F 10.08 7.58e-21 162-180 PR01254E 14.07 1.00e-18 145-159
1068	PR00179	Lipocalin signature II	PR00179B 7.67 5.26e-13 120-132 PR00179C 17.26 3.84e-12 148-163 PR01254B 12.05 9.04e-12 57-67
1068	PR01275	Neutrophil gelatinase lipocalin signature V	PR01275E 6.38 1.72e-10 115-133 PR00179A 13.97 3.25e-10 37-49
1068	PR01215	Alpha-1-microglobulin signature IV	PR01215D 12.88 9.78e-10 111-130
1068	IPB000566	Lipocalin and cytosolic fatty-acid binding protein	IPB000566B 8.91 1.47e-09 120-130
1068	PR01174	Retinol binding protein signature VI	PR01174F 11.76 3.96e-09 119-135
1068	PR01273	Invertebrate colouration protein signature IV	PR01273D 11.48 4.41e-09 120-134 PR01275B 9.02 8.57e-09 39-49
1069	IPB000704	"Casein kinase II, regulatory subunit"	IPB000704B 17.35 6.26e-09 90-128
1070	IPB001464	Annexin family	IPB001464D 25.42 1.00e-40 281-335 IPB001464B 28.31 6.76e-40 151-203 IPB001464A 31.17 1.27e-35 79-133 IPB001464C 24.68 6.40e-30 214-253
1070	PR00196	Annexin family signature IV	PR00196D 21.41 3.81e-22 219-245 PR00196E 9.70 7.75e-21 299-319
1070	PR00201	Annexin type V signature VII	PR00201G 12.46 1.00e-20 299-325 PR00196C 9.01 7.09e-20 136-157 IPB001464B 28.31 4.88e-19 79-131 PR00196A 12.07 2.42e-18 69-91
1070	PR00199	Annexin type III signature VI	PR00199F 15.67 5.10e-18 219-245 IPB001464D 25.42 9.21e-18 122-176 IPB001464B 28.31 3.86e-17 235-287 IPB001464A 31.17 6.68e-17 151-205
1070	PR00200	Annexin type IV signature VII	PR00200G 9.20 8.41e-17 299-325 PR00199D 4.74 2.11e-16 295-316 PR00199G 9.85 5.29e-16 300-325 PR00196C 9.01 5.96e-16 295-316 PR00199D 4.74 7.04e-16 136-157
1070	PR00197	Annexin type I signature IV	PR00197D 7.59 7.56e-16 136-157 PR00196B 11.03 9.31e-16 109-125
1070	PR00198	Annexin type II signature IV	PR00198D 7.41 9.88e-16 136-157 PR00200E 8.88 5.88e-15 136-157 PR00197F 9.40 7.39e-15 299-319
1070	PR00202	Annexin type VI signature VII	PR00202G 8.03 9.71e-15 299-325

405
TABLE 3B

			IPB001464A 31.17 1.85e-14 235-289 PR00197D 7.59 1.94e-14 295-316 PR00196C 9.01 5.02e-14 64-85 PR00201D 8.61 9.29e-14 136-157 PR00199D 4.74 2.84e-13 64-85 PR00198D 7.41 3.15e-13 295-316 PR00
1071	IPB000175	Sodium:neurotransmitter symporter family	IPB000175A 16.29 1.00e-40 52-101 IPB000175C 15.09 1.00e-40 212-263 IPB000175F 25.63 4.50e-38 467-506 IPB000175E 21.88 5.95e-35 372-411 IPB000175B 19.12 9.05e-33 139-173
1071	PR00176	Sodium/chloride neurotransmitter symporter signature I	PR00176A 16.97 3.25e-27 52-73 PR00176C 10.57 7.86e-25 124-150
1071	PR01195	GAT-1 GABA neurotransmitter transporter signature II	PR01195B 13.58 1.22e-24 194-211 PR00176G 13.12 3.77e-22 458-478 PR01195D 9.00 3.75e-21 583-600 PR00176E 11.14 5.20e-21 322-342 PR00176F 11.11 1.36e-19 376-395 IPB000175G 16.18 5.13e-19 528-550 PR00176B 7.07 9.63e-19 81-100 PR01195A 7.44 1.90e-18 18-32 PR00176D 8.96 6.48e-18 239-256 PR00176H 15.94 7.63e-18 498-518 IPB000175D 23.45 1.28e-17 278-330 PR01195C 15.62 1.14e-13 348-357
1072	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.92e-10 98-135
1073	IPB001863	Glypican	IPB001863D 26.43 5.62e-33 250-294 IPB001863E 33.79 3.08e-29 298-350 IPB001863B 38.78 1.45e-25 134-186 IPB001863F 26.99 6.59e-22 429-463 IPB001863C 20.17 1.37e-16 191-220 IPB001863A 13.95 5.03e-15 56-71 IPB001863G 11.32 4.68e-12 487-505
1073	PR00436	Interleukin-8 signature I	PR00436A 15.20 7.91e-10 1-24
1073	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 3.90e-09 515-529
1073	IPB001702	General diffusion Gram-negative porins	IPB001702D 9.64 1.00e-08 536-546
1075	IPB001675	Glycosyltransferase family 29	IPB001675A 26.48 5.76e-31 296-340 IPB001675B 15.84 6.50e-15 434-456
1075	PR01329	Kir3.3 inward rectifier K ⁺ channel signature II	PR01329B 8.30 9.29e-09 7-21
1078	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 7.84e-26 1244-1271 IPB001599F 18.95 7.00e-24 785-814 IPB001599H 18.42 6.40e-20 1019-1046 IPB001599A 10.97 9.69e-18 123-141 IPB001599N 24.85 2.24e-14 1437-1469
1078	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 1257-1271 IPB001599M 13.29 4.71e-13 1384-1395 IPB001599G 13.87 8.94e-13 987-996 IPB001599B 7.45 4.89e-12 209-221 IPB001599D 11.61 6.90e-12 728-738 IPB001599J 20.99 3.00e-11 1085-1110 IPB001599I 10.83 7.60e-11 1054-1063 IPB001599K 8.15 1.46e-10 1214-1225 IPB001599C 14.40 3.55e-09 236-252 IPB001599E 11.06 9.77e-09 755-764
1079	IPB001599	Alpha-2-macroglobulin family	IPB001599F 18.95 7.00e-24 799-828

406

TABLE 3B

			IPB001599A 10.97 9.69e-18 136-154 IPB001599B 7.45 4.89e-12 222-234 IPB001599D 11.61 6.90e-12 742-752 IPB001599C 14.40 3.55e-09 249-265 IPB001599E 11.06 9.77e-09 769-778
1080	IPB001599	Alpha-2-macroglobulin family	IPB001599A 10.97 9.69e-18 123-141 IPB001599B 7.45 4.89e-12 209-221 IPB001599C 14.40 3.55e-09 236-252
1081	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 7.84e-26 1244-1271 IPB001599F 18.95 7.00e-24 785-814 IPB001599H 18.42 6.40e-20 1019-1046 IPB001599N 24.85 7.69e-20 1437-1469 IPB001599A 10.97 9.69e-18 123-141
1081	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 1257-1271 IPB001599M 13.29 4.71e-13 1384-1395 IPB001599G 13.87 8.94e-13 987-996 IPB001599B 7.45 4.89e-12 209-221 IPB001599D 11.61 6.90e-12 728-738 IPB001599J 20.99 3.00e-11 1085-1110 IPB001599I 10.83 7.60e-11 1054-1063 IPB001599K 8.15 1.46e-10 1214-1225 IPB001599C 14.40 3.55e-09 236-252 IPB001599E 11.06 9.77e-09 755-764
1082	IPB001599	Alpha-2-macroglobulin family	IPB001599F 18.95 7.00e-24 786-815 IPB001599A 10.97 9.69e-18 123-141 IPB001599B 7.45 4.89e-12 209-221 IPB001599D 11.61 6.90e-12 729-739 IPB001599C 14.40 3.55e-09 236-252 IPB001599E 11.06 9.77e-09 756-765
1083	IPB002018	Carboxylesterases type-B	IPB002018 21.41 2.38e-27 195-235 IPB002018 21.41 2.47e-12 504-544
1083	PR00878	Cholinesterase signature VI	PR00878F 4.95 8.07e-09 460-472
1084	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 1.64e-16 1682-1697 IPB000152 8.86 1.53e-15 1178-1193 IPB000152 8.86 1.47e-14 1136-1151 IPB000152 8.86 2.89e-14 1095-1110 IPB000152 8.86 3.84e-14 932-947 IPB000152 8.86 4.79e-14 1219-1234 IPB000152 8.86 5.74e-14 642-657 IPB000152 8.86 3.05e-13 1054-1069
1084	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 4.00e-13 1682-1693
1084	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 7.72e-13 1023-1043 IPB001881B 12.28 7.75e-13 1095-1106 IPB000152 8.86 9.18e-13 1261-1276 IPB001881B 12.28 1.00e-12 642-653 IPB001881B 12.28 2.20e-12 1483-1494 IPB000152 8.86 6.40e-12 1483-1498 IPB001881B 12.28 6.40e-12 1178-1189 IPB001881B 12.28 8.20e-12 1261-1272 IPB001881B 12.28 9.40e-12 1136-1147
1084	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 1.00e-11 1136-1155
1084	PR00010	Type II EGF-like signature III	PR00010C 6.98 1.37e-11 1687-1697 IPB001881B 12.28 3.84e-11 1219-1230 PR00010C 6.98 4.00e-11 1183-1193
1084	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 4.24e-11 1059-1069 IPB001881B 12.28 6.68e-11 932-943 IPB003886D 13.91 2.92e-10 1219-1238
1084	IPB003306	WIF domain	IPB003306E 25.51 4.00e-10 176-221

TABLE 3B

1084	IPB000034	Laminin B	IPB000034A 22.21 4.62e-10 187-222 IPB001881B 12.28 5.29e-10 1054-1065 IPB000152 8.86 5.50e-10 1303-1318 IPB000033B 7.05 5.65e-10 1266-1276 IPB000033B 7.05 6.23e-10 1100-1110 IPB001881B 12.28 6.57e-10 1303-1314 IPB001881B 12.28 7.43e-10 1014-1025 IPB000152 8.86 7.75e-10 890-905 IPB000033B 7.05 8.26e-10 1687-1697 PR00010C 6.98 8.55e-10 937-947
1084	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 8.94e-10 175-220 IPB003886D 13.91 1.00e-09 1682-1701 IPB000033B 7.05 1.24e-09 647-657 IPB000033B 7.05 1.47e-09 1141-1151 IPB000033B 7.05 1.95e-09 1183-1193 IPB003306D 23.91 2.18e-09 194-242 PR00010C 6.98 2.32e-09 647-657 IPB003886D 13.91 2.52e-09 1178-1197
1084	PR00011	Type III EGF-like signature IV	PR00011D 12.12 4.21e-09 413-431 IPB003886D 13.91 4.32e-09 1095-1114 IPB001881B 12.28 4.52e-09 890-901 IPB000033B 7.05 4.79e-09 937-947 PR00010C 6.98 4.95e-09 1059-1069 PR00010C 6.98 5.39e-09 1224-1234 IPB000034A 22.21 5.89e-09 399-434 PR00010C 6.98 6.71e-09 1266-1276 IPB001881B 12.28 6.87e-09 1442-1453 IPB000033B 7.05 6.92e-09 1224-1234 IPB003886D 13.91 7.09e-09 1261-1280
1084	IPB002221	WAP-type (Whey Acidic Protein) four-disulfide core domain	IPB002221B 17.12 7.75e-09 1466-1487
1084	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 8.02e-09 92-106
1084	PR00009	Type I EGF signature III	PR00009C 11.70 8.20e-09 1058-1069 IPB000152 8.86 8.58e-09 1637-1652
1084	IPB002557	Chitin binding domain	IPB002557B 12.64 9.31e-09 1453-1466
1084	IPB000561	EGF-like domain	IPB000561 4.89 9.36e-09 1187-1195
1084	IPB002919	Trypsin Inhibitor-like cysteine rich domain	IPB002919B 21.14 9.51e-09 899-921 IPB000152 8.86 9.76e-09 1442-1457 IPB003886D 13.91 9.86e-09 642-661 IPB003886D 13.91 9.86e-09 932-951 IPB000561 4.89 1.00e-08 420-428 PR00010C 6.98 1.00e-08 1141-1151
1086	PR00014	Fibronectin type III repeat signature IV	PR00014D 15.12 9.25e-13 571-585 PR00014C 14.47 6.63e-11 651-669 PR00014D 15.12 7.75e-11 872-886 PR00014D 15.12 5.74e-10 443-457 PR00014C 14.47 6.50e-10 854-872 PR00014A 8.22 1.00e-08 816-825 PR00014D 15.12 1.00e-08 770-784
1087	IPB001909	KRAB box	IPB001909 17.37 7.75e-31 16-50
1087	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.55e-21 219-244 IPB000822 14.67 4.21e-17 191-216 IPB000822 14.67 8.80e-16 163-188
1087	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 5.85e-14 188-201 PR00048A 9.94 9.31e-14 244-257 PR00048A 9.94 8.41e-12 216-229
1087	IPB001275	DM DNA binding domain	IPB001275 19.17 5.24e-11 207-246 PR00048A 9.94 7.16e-11 160-173

408

TABLE 3B

			PR00048B 5.52 6.14e-10 232-241 IPB001275 19.17 7.45e-10 151-190 IPB001275 19.17 8.06e-09 179-218
1088	IPB001909	KRAB box	IPB001909 17.37 7.75e-31 16-50
1088	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 6.21e-11 160-173
1089	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 8.36e-35 20-63 IPB002494C 14.46 5.74e-34 89-132 IPB002494C 14.46 1.44e-30 99-142 IPB002494C 14.46 7.86e-29 64-107 IPB002494C 14.46 1.41e-27 74-117 IPB002494C 14.46 4.71e-25 30-73 IPB002494C 14.46 6.69e-25 79-
1089	IPB000359	Cystine-knot domain	IPB000359B 19.26 9.57e-13 24-42 IPB000359B 19.26 9.57e-13 68-86 IPB002494C 14.46 9.61e-13 73-116 IPB002494B 10.58 2.50e-12 51-65 IPB002494B 10.58 2.50e-12 95-109 IPB002494C 14.46 4.37e-12 34-77 IPB002494A 12.44 5.22e-12 91-124 IPB002494C 14.46 6.06e-12 93-136 IPB002494C 14.46 7.47e-12 83-126
1089	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 7.62e-12 66-111 IPB002494B 10.58 7.75e-12 65-79
1089	IPB001271	Mammalian defensin	IPB001271 19.97 7.95e-12 58-86 IPB002494B 10.58 9.55e-12 120-134 IPB001271 19.97 9.59e-12 19-47 IPB002494B 10.58 1.28e-11 26-40 IPB002494B 10.58 1.28e-11 70-84 IPB002494A 12.44 1.86e-11 121-154 IPB002494A 12.44 2.82e-11 56-89 IPB001271 19.97 3.06e-11 103-131 IPB000006 13.41 4.50e-11 70-115 IPB000006 13.41 5.50e-11 40-85 IPB002494C 14.46 6.64e-11 98-141 IPB002494C 14.46 6.73e-11 78-121 IPB000006 13.41 8.20e-11 65-110 IPB002494A 12.44 9.14e-11 57-90 IPB001271 19.97 1.88e-10 28-56 IPB001271 19.97 1.88e-10 72-100 IPB002494C 14.46 2.14e-10 14-57 IPB002494B 10.58 2.48e-10 56-70 IPB000006 13.41 2.65e-10 61-106 IPB001271 19.97 2.94e-10 67-95 IPB001271 19.97 3.12e-10 18-46 IPB000006 13.41 3.42e-10 22-67 IPB002494B 10.58 4.22e-10 110-124
1089	IPB001762	Disintegrin	IPB001762A 23.93 4.26e-10 39-79 IPB002494A 12.44 4.27e-10 46-79 IPB000006 13.41 4.29e-10 21-66 IPB001762A 23.93 4.45e-10 44-84 IPB001271 19.97 5.41e-10 117-145 IPB000006 13.41 6.23e-10 91-136 IPB001271 19.97 6.47e-10 123-151 IPB000006 13.41 6.61e-10 26-71 IPB002494B 10.58 6.64e-10 31-45 IPB002494B 10.58 6.64e-10 75-89 IPB002494B 10.58 6.91e-10 41-55 IPB002494B 10.58 6.91e-10 85-99

409

TABLE 3B

			IPB002494C 14.46 7.64e-10 108-151 IPB002494A 12.44 7.65e-10 67-100 IPB002494B 10.58 7.72e-10 100-114 IPB002494A 12.44 8.06e-10 82-115 IPB002494C 14.46 8.25e-10 19-62
1089	IPB000967	Zinc finger NF-X1 type	IPB000967E 21.88 8.67e-10 51-91 IPB000359B 19.26 8.76e-10 59-77 IPB001271 19.97 8.76e-10 88-116 IPB000006 13.41 9.03e-10 114-159 IPB001762A 23.93 9.04e-10 45-85 IPB001762A 23.93 9.04e-10 94-134 IPB002494C 14.46 9.48e-10 4-47
1089	IPB001169	"Integrin beta, C-terminus"	IPB001169K 27.45 4.89e-09 86-128 IPB001271 19.97 4.93e-09 29-57 IPB001271 19.97 4.93e-09 73-101 IPB001271 19.97 4.93e-09 97-125 IPB001271 19.97 4.93e-09 102-130 IPB002494C 14.46 4.95e-09 65-108 IPB000006 13.41 5.22e-09 81-126
1090	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 9.43e-29 24-67 IPB002494C 14.46 3.22e-22 14-57 IPB002494C 14.46 8.08e-21 29-72 IPB002494C 14.46 7.99e-20 19-62 IPB002494A 12.44 3.29e-19 31-64 IPB002494C 14.46 8.65e-18 9-52 IPB002494A 12.44 8.15e-17 21-54 IPB002494A 12.44 7.17e-16 36-69 IPB002494A 12.44 6.12e-15 2-35 IPB002494A 12.44 4.96e-14 26-59 IPB002494C 14.46 2.86e-13 5-48 IPB002494C 14.46 4.72e-13 28-71 IPB002494C 14.46 5.30e-13 4-47 IPB002494A 12.44 6.19e-13 12-45 IPB002494A 12.44 6.54e-13 41-74 IPB002494A 12.44 8.15e-13 1-34 IPB002494C 14.46 9.51e-13 20-63
1090	IPB000359	Cystine-knot domain	IPB000359B 19.26 9.57e-13 28-46
1090	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 4.21e-12 26-71
1090	IPB001271	Mammalian defensin	IPB001271 19.97 7.75e-12 18-46 IPB002494A 12.44 1.11e-11 11-44 IPB002494B 10.58 1.28e-11 30-44 IPB002494A 12.44 6.25e-11 16-49 IPB002494C 14.46 8.27e-11 15-58 IPB002494A 12.44 8.39e-11 6-39 IPB002494C 14.46 9.82e-11 10-53
1090	IPB001762	Disintegrin	IPB001762A 23.93 9.65e-09 34-74 IPB002494A 12.44 9.90e-09 27-60 IPB000006 13.41 1.00e-08 25-70
1091	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 8.36e-35 20-63 IPB002494C 14.46 7.86e-32 124-167 IPB002494C 14.46 6.55e-31 64-107 IPB002494C 14.46 8.95e-31 89-132 IPB002494C 14.46 1.44e-30 134-177 IPB002494C 14.46 4.23e-28 99-142 IPB002494C 14.46 9.46e-26
1091	IPB000359	Cystine-knot domain	IPB000359B 19.26 9.57e-13 24-42 IPB000359B 19.26 9.57e-13 68-86

410

TABLE 3B

			IPB002494A 12.44 1.56e-12 42-75 IPB002494B 10.58 2.50e-12 51-65 IPB002494B 10.58 2.50e-12 95-109 IPB002494B 10.58 2.50e-12 130-144 IPB002494C 14.46 5.41e-12 34-77 IPB002494C 14.46 6.06e-12 128-171 IPB002494C 14.46 7.28e-12 118-161
1091	IPB001271	Mammalian defensin	IPB001271 19.97 7.95e-12 58-86 IPB002494C 14.46 9.25e-12 103-146 IPB002494B 10.58 9.55e-12 155-169 IPB001271 19.97 9.59e-12 19-47 IPB002494B 10.58 1.28e-11 26-40 IPB002494B 10.58 1.28e-11 70-84 IPB002494A 12.44 1.86e-11 156-189 IPB001271 19.97 3.06e-11 138-166 IPB002494A 12.44 4.00e-11 56-89
1091	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 4.10e-11 66-111 IPB002494C 14.46 4.91e-11 113-156 IPB001271 19.97 5.13e-11 97-125 IPB002494C 14.46 6.64e-11 133-176 IPB000006 13.41 6.80e-11 40-85 IPB000359B 19.26 7.48e-11 103-121 IPB002494C 14.46 7.91e-11 98
1091	IPB001762	Disintegrin	IPB001762A 23.93 9.04e-10 129-169 IPB002494C 14.46 9.21e-10 65-108 IPB000006 13.41 9.42e-10 95-140 IPB002494C 14.46 9.48e-10 4-47 IPB000359B 19.26 9.69e-10 158-176 IPB000359B 19.26 1.28e-09 153-171 IPB000006 13.41 1.55e-09 115-160
1091	IPB000967	Zinc finger NF-X1 type	IPB000967E 21.88 1.56e-09 51-91 IPB002494A 12.44 1.58e-09 147-180 IPB001762A 23.93 1.88e-09 39-79 IPB001271 19.97 2.15e-09 98-126 IPB002494A 12.44 2.55e-09 62-95 IPB002494A 12.44 3.13e-09 41-74 IPB002494A 12.44 3.23e-09 28-61 IPB002494A 12.44 3.23e-09 72-105 IPB002494A 12.44 3.23e-09 77-110 IPB002494B 10.58 3.41e-09 16-30 IPB001271 19.97 3.78e-09 23-51 IPB001271 19.97 3.78e-09 67-95
1091	IPB001169	"Integrin beta, C-terminus"	IPB001169K 27.45 3.92e-09 121-163 IPB000006 13.41 3.94e-09 80-125 IPB000006 13.41 4.03e-09 140-185 IPB001762A 23.93 4.18e-09 44-84 IPB002494B 10.58 4.42e-09 125-139 IPB002494A 12.44 4.48e-09 33-66 IPB000006 13.41 4.86e-09 65-
1092	IPB000734	Lipase	IPB000734 10.25 8.12e-09 164-178
1093	IPB000734	Lipase	IPB000734 10.25 8.12e-09 224-238
1094	PR01223	Bride of sevenless protein signature VI	PR01223F 4.19 9.78e-11 203-227
1094	PR00354	7Fe ferredoxin signature III	PR00354C 6.24 8.06e-09 258-275
1096	IPB001304	C-type lectin domain	IPB001304A 17.98 8.04e-14 87-111
1096	PR00356	Type II antifreeze protein signature VII	PR00356G 10.21 1.42e-10 193-206
1097	IPB001304	C-type lectin domain	IPB001304A 17.98 8.04e-14 87-111

411

TABLE 3B

1097	PR00356	Type II antifreeze protein signature VII	PR00356G 10.21 8.15e-09 193-206
1098	PR00245	Olfactory receptor signature V	PR00245E 8.96 5.15e-16 283-294 PR00245B 13.73 3.77e-15 129-141 PR00245C 14.65 2.73e-14 176-192 PR00245D 9.34 2.59e-13 236-245
1098	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 7.00e-12 118-129 PR00245A 10.98 1.72e-11 92-103 IPB000276D 9.40 6.09e-10 282-298
1098	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 2.83e-09 51-63
1098	PR00237	Rhodopsin-like GPCR superfamily signature III	PR00237C 14.77 3.86e-09 104-126 PR00237B 12.45 6.92e-09 59-80 PR00237A 9.81 8.31e-09 26-50
1099	IPB002889	WSC domain	IPB002889B 11.76 3.44e-09 56-102
1099	IPB000561	EGF-like domain	IPB000561 4.89 4.86e-09 306-314
1099	IPB000034	Laminin B	IPB000034C 12.97 7.43e-09 306-324
1099	PR00346	Tissue factor signature VIII	PR00346H 10.74 8.18e-09 542-565
1101	PR00457	Animal haem peroxidase signature V	PR00457E 19.97 8.45e-24 997-1023 PR00457D 18.35 1.53e-20 972-992 PR00457C 18.81 9.42e-15 954-972 PR00457G 14.17 4.48e-14 1177-1197 PR00457H 14.82 5.85e-13 1248-1262 PR00457F 14.42 6.32e-12 1050-1060
1101	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 1.00e-10 180-194 PR00457B 12.43 2.29e-10 802-817
1101	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 2.80e-10 376-413 IPB003006B 20.23 8.92e-10 466-503 IPB003006B 20.23 9.28e-10 283-320
1101	PR00019	Leucine-rich repeat signature II	PR00019B 11.42 6.73e-09 73-86
1102	PR00457	Animal haem peroxidase signature V	PR00457E 19.97 8.45e-24 973-999 PR00457D 18.35 1.53e-20 948-968 PR00457C 18.81 9.42e-15 930-948 PR00457G 14.17 4.48e-14 1153-1173 PR00457H 14.82 5.85e-13 1224-1238 PR00457F 14.42 6.32e-12 1026-1036
1102	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 1.00e-10 156-170 PR00457B 12.43 2.29e-10 778-793
1102	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 2.80e-10 352-389 IPB003006B 20.23 8.92e-10 442-479 IPB003006B 20.23 9.28e-10 259-296
1103	IPB002034	Alpha-isopropylmalate and homocitrate synthase	IPB002034D 19.67 7.61e-09 786-814
1107	IPB001359	Synapsin	IPB001359H 22.58 1.80e-14 741-791
1107	IPB000885	Fibrillar collagen C-terminal domain	IPB000885A 11.46 8.16e-09 765-802
1107	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 9.14e-09 746-798
1110	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 3.52e-10 31-68
1112	IPB001841	RING finger	IPB001841 10.69 1.95e-09 153-162
1113	IPB000961	Protein kinase C-terminal domain	IPB000961A 16.82 2.64e-12 193-227
1113	IPB000959	POLO box duplicated region	IPB000959B 15.68 9.22e-12 288-328
1113	IPB001245	Tyrosine kinase catalytic domain	IPB001245A 22.45 1.87e-11 304-344
1113	IPB001772	Kinase associated domain 1	IPB001772C 20.66 6.11e-11 299-329
1113	IPB003527	MAP kinase	IPB003527C 14.70 3.43e-09 296-344
1119	PR01137	Gap junction alpha-8 protein (Cx50) signature II	PR01137B 18.37 8.83e-09 368-380

412
TABLE 3B

1120	IPB000906	ZU5 domain	IPB000906G 25.85 2.58e-13 921-969 IPB000906F 35.93 9.00e-12 931-984 IPB000906D 23.89 1.57e-11 940-994
1120	PR00452	SH3 domain signature II	PR00452B 11.47 2.73e-11 1036-1051
1120	PR01415	Ankyrin repeat signature I	PR01415A 12.73 6.46e-11 954-966 IPB000906A 22.49 7.53e-10 914-956 PR01415A 12.73 7.97e-10 921-933
1120	PR00499	Neutrophil cytosol factor 2 signature IV	PR00499D 11.47 4.21e-09 1024-1044
1120	IPB002360	Involucrin	IPB002360C 15.36 4.90e-09 125-166 IPB000906F 35.93 7.41e-09 898-951
1120	IPB000237	GRIP domain	IPB000237B 30.66 8.14e-09 142-192
1124	IPB000906	ZU5 domain	IPB000906D 23.89 7.66e-10 117-171 IPB000906A 22.49 3.72e-09 58-100 IPB000906G 25.85 6.69e-09 164-212
1125	IPB000906	ZU5 domain	IPB000906D 23.89 7.66e-10 117-171 IPB000906A 22.49 3.72e-09 58-100
1129	IPB000421	Coagulation factor 5/8 type C domain (FA58C)	IPB000421C 36.74 1.93e-16 131-175 IPB000421B 20.70 1.36e-14 79-99
1130	IPB000421	Coagulation factor 5/8 type C domain (FA58C)	IPB000421C 36.74 1.93e-16 131-175 IPB000421B 20.70 1.36e-14 79-99
1130	PR01435	NADH-plastoquinone oxidoreductase chain 5 signature II	PR01435B 5.98 7.37e-10 1059-1083
1131	IPB002119	Histone H2A	IPB002119A 4.97 1.00e-08 92-98
1133	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 4.43e-18 178-216
1133	IPB003527	MAP kinase	IPB003527D 21.53 3.41e-16 171-212
1133	IPB000961	Protein kinase C-terminal domain	IPB000961A 16.82 6.56e-15 10-44
1133	IPB000861	PKN/rhopilin/rhotekin rho-binding repeat	IPB000861D 13.61 6.92e-15 8-44
1133	IPB000959	POLO box duplicated region	IPB000959C 23.49 6.34e-14 153-205 IPB003527G 17.26 4.28e-13 320-357 IPB001245A 22.45 8.07e-13 119-159
1133	IPB001772	Kinase associated domain 1	IPB001772C 20.66 4.51e-12 114-144 IPB000861G 13.73 5.06e-12 180-229
1133	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 1.18e-11 182-236 IPB000961D 21.23 1.00e-10 174-215 IPB001772A 13.64 1.86e-10 8-39 IPB003527A 17.00 2.75e-10 17-42 IPB000959B 15.68 9.10e-10 103-143
1135	PR00402	Tec/Btk domain signature I	PR00402A 20.14 8.15e-15 664-683 PR00402B 12.26 4.69e-13 683-695
1135	PR00360	C2 domain signature II	PR00360B 11.64 9.25e-13 174-187 PR00402C 13.13 8.03e-12 695-708
1135	IPB000008	C2 domain	IPB000008D 14.83 1.61e-11 200-218 PR00360A 15.18 6.00e-10 150-162 PR00360A 15.18 8.33e-10 22-34
1135	PR00399	Synaptotagmin signature IV	PR00399D 12.72 4.89e-09 79-89 PR00360C 7.35 5.50e-09 196-204
1137	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.57e-15 261-280
1137	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 7.16e-14 134-149 IPB000152 8.86 9.05e-14 216-231 IPB000152 8.86 5.91e-13 261-276
1137	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 9.25e-13 216-227
1137	IPB001774	Delta serrate ligand	IPB001774C 18.25 9.69e-12 66-108 IPB001881B 12.28 1.95e-11 134-145
1137	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 4.96e-11 266-276

413

TABLE 3B

1137	PR01217	Proline rich extensin signature VII	PR01217G 4.02 5.15e-11 340-365
1137	PR00907	Thrombomodulin signature II	PR00907B 11.50 6.70e-11 168-184 IPB001881B 12.28 1.00e-10 261-272
1137	IPB000925	Pneumovirus attachment glycoprotein G	IPB000925F 15.07 3.60e-10 336-372
1137	IPB000561	EGF-like domain	IPB000561 4.89 6.25e-10 75-83
1137	PR00010	Type II EGF-like signature III	PR00010C 6.98 1.66e-09 266-276
1137	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 3.29e-09 348-362 PR00049D 0.00 3.29e-09 350-364 IPB000033B 7.05 3.84e-09 221-231 PR01217E 3.04 4.48e-09 348-364 PR01217B 4.82 6.55e-09 347-363 IPB000561 4.89 6.79e-09 270-278 PR00010C 6.98 7.15e-09 139-149 PR01217D 4.57 7.16e-09 343-364 PR00010C 6.98 7.80e-09 221-231 IPB000033B 7.05 8.11e-09 139-149
1137	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 8.62e-09 183-203
1137	PR00910	Luteovirus ORF6 protein signature I	PR00910A 2.74 8.71e-09 348-360 PR00910A 2.74 9.46e-09 346-358 PR01217G 4.02 9.92e-09 343-368
1138	IPB001156	Transferrin	IPB001156H 23.81 7.75e-09 118-172
1143	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 59-75
1143	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 1-12 PR00245D 9.34 1.53e-13 119-128 PR00245E 8.96 6.81e-12 166-177 PR00245B 13.73 1.00e-10 12-24 IPB000276D 9.40 3.08e-09 165-181
1143	PR00237	Rhodopsin-like GPCR superfamily signature V	PR00237E 13.03 3.83e-09 82-105 PR00237G 19.23 1.00e-08 155-181
1144	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 173-189
1144	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 117-128 PR00245D 9.34 1.53e-13 233-242 PR00245E 8.96 6.81e-12 280-291 PR00245A 10.98 7.14e-12 91-102 PR00245B 13.73 8.14e-10 128-140
1144	PR00237	Rhodopsin-like GPCR superfamily signature III	PR00237C 14.77 2.02e-09 103-125 IPB000276D 9.40 3.08e-09 279-295 PR00237E 13.03 3.83e-09 196-219
1144	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.17e-09 50-62
1144	PR00896	Vasopressin receptor signature II	PR00896B 9.36 7.23e-09 54-65 PR00237G 19.23 1.00e-08 269-295
1146	IPB000017	Syntaxin / epimorphin family	IPB000017 23.80 1.84e-09 168-217
1147	PR01360	Interleukin-1 receptor antagonist precursor IL-1RA signature VI	PR01360F 14.44 3.11e-12 117-135 PR01360C 10.33 4.84e-11 58-75
1147	IPB000975	Interleukin-1	IPB000975D 24.45 5.55e-09 52-91 IPB000975E 28.12 9.80e-09 96-135
1147	PR00264	Interleukin-1 precursor family signature I	PR00264A 18.63 1.00e-08 55-75
1148	PR01248	Type I keratin signature V	PR01248E 12.72 3.67e-21 248-274
1148	IPB001664	Intermediate filament proteins	IPB001664B 17.44 9.16e-20 104-143 IPB001664A 11.94 8.13e-19 381-406 PR01248C 10.07 8.34e-17 150-170
1148	IPB001322	Intermediate filament tail domain	IPB001322A 30.52 2.23e-14 370-423 IPB001664C 11.32 3.25e-13 161-188 PR01248B 8.42 3.29e-13 96-119

414

TABLE 3B

			PR01248D 9.34 3.60e-12 222-237 PR01248A 8.12 6.14e-11 75-88
1148	PR01177	Metabotropic gamma-aminobutyric acid type B1 receptor signature X	PR01177J 6.10 4.96e-10 397-415 PR01177J 6.10 3.63e-09 13-31 IPB001664D 12.63 5.36e-09 279-305
1151	IPB001664	Intermediate filament proteins	IPB001664D 12.63 4.75e-28 384-410
1151	PR01276	Type II keratin signature IV	PR01276D 13.08 8.31e-24 222-241 IPB001664A 11.94 9.50e-23 132-157
1151	IPB001322	Intermediate filament tail domain	IPB001322C 22.70 4.75e-22 374-419 IPB001664C 11.32 8.20e-21 266-293 PR01276E 12.04 4.75e-15 301-318 IPB001322A 30.52 4.08e-14 121-174 PR01276F 10.92 3.21e-11 352-367 PR01276C 10.16 8.66e-11 208-221 IPB001664B 17.44 5.27e-10 191-230 PR01276B 9.79 5.96e-10 161-173 PR01276A 10.31 7.16e-10 134-142
1151	IPB003743	DUF164	IPB003743B 20.16 9.21e-10 300-338
1152	IPB001818	Matrixin	IPB001818C 24.38 8.03e-32 157-202 IPB001818B 26.48 6.04e-31 112-153 IPB001818A 14.60 2.13e-29 66-95 IPB001818H 15.46 3.25e-23 332-358 IPB001818F 11.19 4.91e-20 231-251
1152	PR00138	Matrixin signature I	PR00138A 12.54 1.64e-16 86-99 PR00138C 20.07 1.78e-16 155-183 IPB001818G 14.71 1.96e-12 268-280 PR00138B 14.84 5.21e-10 131-146
1153	IPB001818	Matrixin	IPB001818C 24.38 8.03e-32 157-202 IPB001818B 26.48 6.04e-31 112-153 IPB001818A 14.60 2.13e-29 66-95 IPB001818H 15.46 3.25e-23 332-358 IPB001818F 11.19 4.91e-20 231-251
1153	PR00138	Matrixin signature I	PR00138A 12.54 1.64e-16 86-99 PR00138C 20.07 1.78e-16 155-183 IPB001818G 14.71 1.96e-12 268-280 PR00138B 14.84 5.21e-10 131-146
1154	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 2.07e-09 10-24
1154	IPB002000	Lysosome-associated membrane glycoprotein (Lamp)	IPB002000D 5.87 5.25e-09 12-25
1155	IPB001124	Lipid-binding serum glycoprotein	IPB001124C 25.71 7.71e-17 210-253 IPB001124D 21.85 5.71e-14 274-310
1156	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135B 13.24 9.39e-10 84-128 IPB000135A 11.69 6.19e-09 111-165
1156	IPB003533	Doublecortin	IPB003533H 6.52 7.51e-09 49-72
1159	IPB001510	Poly(ADP-ribose) polymerase zinc finger domain	IPB001510D 30.92 1.00e-40 490-543 IPB001510E 22.53 1.00e-40 570-624 IPB001510A 34.80 7.21e-40 92-137 IPB001510B 23.09 6.14e-34 306-348 IPB001510C 15.91 6.54e-27 363-396
1159	IPB000977	ATP-dependent DNA ligase	IPB000977B 14.05 4.60e-13 508-517 IPB000977C 7.51 1.00e-12 590-599 IPB000977A 8.89 1.47e-09 480-487
1160	IPB000215	Serpins	IPB000215E 15.36 5.50e-23 401-425 IPB000215D 15.35 6.82e-21 317-343 IPB000215A 13.01 7.43e-18 27-50 IPB000215C 13.90 3.16e-12 207-221 IPB000215B 9.87 9.59e-11 178-190

415

TABLE 3B

1166	IPB001309	ICE-like protease (caspase) p20 domain	IPB001309A 10.71 3.57e-14 7-17
1166	PR00376	Interleukin-1B converting enzyme signature I	PR00376A 12.81 1.61e-10 5-18
1168	IPB000364	Phosphoenolpyruvate carboxykinase (GTP)	IPB000364M 26.08 1.40e-09 589-623
1169	IPB001304	C-type lectin domain	IPB001304A 17.98 6.50e-17 118-142
1171	PR00320	G protein beta WD-40 repeat signature II	PR00320B 12.82 6.62e-13 478-492
1171	PR00308	Type I antifreeze protein signature I	PR00308A 3.72 8.17e-13 158-172 PR00320A 13.15 2.89e-12 478-492 PR00320C 12.32 4.18e-12 247-261 PR00320C 12.32 4.71e-12 478-492 PR00320B 12.82 7.75e-12 247-261 PR00320A 13.15 8.11e-12 427-441 PR00320A 13.15 9.05e-12 247-261 PR00308B 3.38 9.27e-12 161-172 PR00308A 3.72 9.76e-12 162-176 PR00308C 2.79 1.00e-11 161-170
1171	PR01511	Kv1.4 voltage-gated K ⁺ channel signature IV	PR01511D 3.91 3.02e-11 163-173 PR00320C 12.32 3.57e-11 427-441 PR00320B 12.82 5.09e-11 520-534 PR00320B 12.82 7.14e-11 427-441 PR00320A 13.15 7.55e-11 520-534 PR00320C 12.32 4.52e-10 520-534
1171	PR00833	Pollen allergen Poa pI signature VIII	PR00833H 2.61 8.56e-10 164-178 PR00308C 2.79 8.77e-10 165-174 PR01511D 3.91 9.88e-10 159-169
1171	IPB001680	G-protein beta WD-40 repeats	IPB001680 10.43 1.45e-09 429-440 PR00308B 3.38 1.76e-09 165-176 IPB001680 10.43 3.70e-09 480-491 IPB001680 10.43 4.15e-09 249-260
1171	PR00456	Ribosomal protein P2 signature V	PR00456E 3.08 5.08e-09 163-177 PR00308A 3.72 6.74e-09 159-173 PR00320A 13.15 7.75e-09 303-317 PR00833H 2.61 7.78e-09 161-175 PR00320B 12.82 8.45e-09 344-358
1171	IPB000102	Neuraxin / MAP1B repeat	IPB000102A 10.50 8.88e-09 156-184 IPB001680 10.43 9.10e-09 522-533 IPB000102A 10.50 9.22e-09 160-188 PR00308B 3.38 9.75e-09 162-173
1175	IPB001559	Phosphotriesterase family	IPB001559D 19.17 5.00e-20 176-202 IPB001559C 16.25 5.34e-16 141-162 IPB001559E 16.18 5.35e-16 214-232 IPB001559A 10.81 1.23e-11 18-29 IPB001559B 12.98 8.50e-10 122-132
1183	IPB003817	Phosphatidylserine decarboxylase	IPB003817D 23.34 8.71e-25 338-364 IPB003817C 10.66 4.00e-15 316-328 IPB003817E 13.21 2.67e-14 427-443 IPB003817A 12.64 4.15e-13 162-176
1184	IPB000580	TSC-22 / Dip / Bun family	IPB000580 14.33 1.00e-40 116-170
1185	PR00072	Malic enzyme signature IV	PR00072D 12.09 9.29e-09 571-589
1187	PR00901	Pheromone B alpha-1 receptor signature VIII	PR00901H 14.75 4.05e-09 56-66
1188	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469I 10.99 4.86e-16 747-765 IPB002469H 21.17 6.14e-16 702-737 IPB002469J 8.97 3.52e-12 829-845

416
TABLE 3B

1188	IPB002471	Prolyl endopeptidase family serine active site	IPB002471B 24.90 3.66e-11 734-765 IPB002469G 26.76 9.24e-11 657-695
1189	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469I 10.99 4.86e-16 747-765 IPB002469H 21.17 6.14e-16 702-737 IPB002469J 8.97 3.52e-12 791-807
1189	IPB002471	Prolyl endopeptidase family serine active site	IPB002471B 24.90 3.66e-11 734-765 IPB002469G 26.76 9.24e-11 657-695
1190	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469I 10.99 4.86e-16 734-752 IPB002469H 21.17 6.14e-16 689-724 IPB002469J 8.97 3.52e-12 816-832
1190	IPB002471	Prolyl endopeptidase family serine active site	IPB002471B 24.90 3.66e-11 721-752 IPB002469G 26.76 9.24e-11 644-682
1191	IPB000524	"Bacterial regulatory proteins, GntR family"	IPB000524 18.80 7.19e-10 54-94
1193	IPB000906	ZU5 domain	IPB000906A 22.49 6.14e-19 241-283 IPB000906F 35.93 3.09e-16 159-212 IPB000906F 35.93 7.91e-16 192-245
1193	PR01415	Ankyrin repeat signature I	PR01415A 12.73 3.70e-15 348-360 IPB000906A 22.49 1.71e-14 142-184 PR01415A 12.73 9.10e-13 799-811 IPB000906F 35.93 1.00e-12 442-495 IPB000906A 22.49 5.66e-12 208-250 IPB000906G 25.85 9.36e-12 149-197 PR01415A 12.73 1.00e-11 1
1194	PR00834	HtrA/DegQ protease family signature III	PR00834C 15.48 7.35e-19 253-277 PR00834D 11.75 7.39e-17 291-308 PR00834B 10.17 3.25e-13 212-232 PR00834E 13.43 6.03e-12 313-330
1194	IPB000126	"Serine proteases, V8 family"	IPB000126B 12.50 6.81e-12 296-312 PR00834A 8.79 1.44e-11 191-203 PR00834F 11.11 1.53e-09 374-386 IPB000126A 11.75 9.83e-09 183-198
1195	PR00424	Adenosine receptor signature IV	PR00424D 13.35 4.34e-22 21-40
1195	PR00555	Adenosine A3 receptor signature V	PR00555E 7.35 4.75e-21 105-122 PR00555F 11.48 2.74e-20 152-169 PR00555D 10.79 9.36e-19 60-76 PR00424E 14.23 3.75e-14 74-87
1195	PR00237	Rhodopsin-like GPCR superfamily signature VII	PR00237G 19.23 4.21e-14 119-145 PR00237F 14.34 9.28e-14 83-107 PR00237E 13.03 4.60e-12 33-56
1195	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276D 9.40 7.30e-12 129-145 PR00424F 8.75 9.07e-12 119-129
1197	PR00245	Olfactory receptor signature IV	PR00245D 9.34 1.53e-13 241-250 PR00245C 14.65 1.56e-12 181-197
1197	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 5.20e-12 123-134
1197	PR00237	Rhodopsin-like GPCR superfamily signature III	PR00237C 14.77 6.73e-11 109-131 PR00245E 8.96 3.30e-10 288-299 PR00237E 13.03 4.77e-10 204-227 PR00245A 10.98 3.65e-09 97-108 PR00245B 13.73 4.60e-09 134-146
1197	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 8.43e-09 56-68
1198	PR00505	D12 class N6 adenine-specific DNA methyltransferase signature I	PR00505A 15.44 3.67e-12 30-46 PR00505B 11.79 8.88e-12 51-65
1199	PR01254	Prostaglandin D synthase signature I	PR01254A 12.32 6.38e-10 25-48
1199	PR00179	Lipocalin signature II	PR00179B 7.67 2.35e-09 111-123 PR00179A 13.97 5.80e-09 31-43

TABLE 3B

			PR00179C 17.26 6.70e-09 138-153
1199	PR01174	Retinol binding protein signature VI	PR01174F 11.76 6.82e-09 110-126 PR01254E 14.07 8.23e-09 135-149
1199	PR01275	Neutrophil gelatinase lipocalin signature II	PR01275B 9.02 1.00e-08 33-43
1200	PR01042	Aspartyl-tRNA synthetase signature IV	PR01042D 11.70 2.67e-14 432-446 PR01042B 12.76 4.69e-11 233-246 PR01042C 16.81 5.50e-11 393-409 PR01042A 9.01 9.77e-10 217-229
1200	IPB002106	Aminoacyl-transfer RNA synthetases class-II	IPB002106A 13.35 1.00e-08 169-181
1201	PR01217	Proline rich extensin signature VII	PR01217G 4.02 8.03e-09 528-553
1202	IPB003952	Fumarate reductase / succinate dehydrogenase FAD-binding site	IPB003952E 9.04 2.46e-16 31-48
1203	IPB001895	Guanine-nucleotide dissociation stimulators CDC25 family	IPB001895C 20.83 8.50e-23 297-332
1204	IPB000958	KH domain	IPB000958 6.84 5.09e-12 112-125 IPB000958 6.84 2.29e-11 28-41 IPB000958 6.84 7.88e-10 276-289
1207	IPB001393	Calsequestrin	IPB001393A 16.72 1.00e-40 29-78 IPB001393B 11.93 1.00e-40 132-185 IPB001393C 16.33 1.00e-40 188-240 IPB001393D 11.26 1.00e-40 283-335
1207	PR00312	Calsequestrin signature V	PR00312E 8.61 7.75e-36 163-192 PR00312I 15.97 5.71e-35 326-354 PR00312F 16.12 7.87e-35 193-222 PR00312H 13.19 2.80e-34 257-284 PR00312J 13.61 6.48e-34 357-385 PR00312D 9.10 7.17e-33 122-151 PR00312B 14.57 4.41e-32 56-85 PR00312C 16.48 5.62e-32 86-115 PR00312G 11.43 1.49e-31 224-251 PR00312A 11.96 7.94e-27 29-52
1209	IPB002151	Kinesin light chain repeat	IPB002151A 11.63 5.55e-10 275-305
1209	PR00985	Leucyl-tRNA synthetase signature I	PR00985A 10.14 8.25e-09 515-532
1210	IPB000353	"Class II histocompatibility antigen, beta chain, beta-1 domain"	IPB000353B 19.16 7.89e-16 137-186
1210	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006A 17.51 7.63e-15 158-180
1210	IPB001003	"MHC Class II, alpha chain, alpha-1 domain"	IPB001003B 14.72 3.87e-10 145-188
1213	PR00205	Cadherin signature II	PR00205B 20.09 8.31e-23 244-273
1213	IPB002126	Cadherin domain	IPB002126B 12.04 5.80e-16 232-249 PR00205D 12.22 7.26e-15 436-455 PR00205F 19.57 1.64e-14 515-541 PR00205G 13.05 4.86e-14 549-566 PR00205A 17.38 7.88e-14 75-94 PR00205D 12.22 3.40e-13 331-350 PR00205D 12.22 5.80e-13 223-242
1214	IPB001580	Calreticulin family	IPB001580D 12.66 2.71e-38 259-294 IPB001580B 18.74 1.90e-35 166-201
1214	PR00626	Calreticulin signature IV	PR00626D 7.86 9.00e-30 242-264 IPB001580A 12.93 8.71e-28 91-113 PR00626E 10.35 4.68e-23 280-299 PR00626B 14.56 6.06e-20 126-142 PR00626E 10.35 8.00e-19 266-285 PR00626A 14.93 6.50e-18 100-118

418
TABLE 3B

			PR00626C 9.33 8.71e-18 215-228 IPB001580C 9.76 1.56e-17 242-254 IPB001580D 12.66 2.38e-16 245-280 IPB001580D 12.66 8.34e-16 273-308 IPB001580C 9.76 4.30e-15 208-220 IPB001580C 9.76 4.16e-14 225-237 PR00626C 9.33 7.75e-12 232-245 PR00626D 7.86 9.14e-09 208-230
1215	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 3.90e-12 32-77 IPB000006 13.41 4.41e-12 39-84 IPB000006 13.41 6.70e-11 35-80
1215	PR01228	Eggshell protein signature III	PR01228C 5.69 1.22e-10 26-41 PR01228C 5.69 1.98e-10 10-25
1215	IPB001271	Mammalian defensin	IPB001271 19.97 3.29e-10 51-79
1215	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 3.36e-10 45-88 IPB001271 19.97 3.47e-10 29-57 IPB002494A 12.44 6.11e-10 70-103
1215	IPB002174	Furin-like cysteine rich region	IPB002174A 30.51 7.32e-10 11-42 IPB002174A 30.51 7.81e-10 3-34 PR01228C 5.69 8.05e-10 19-34
1215	IPB003571	Snake toxin	IPB003571B 18.08 8.07e-10 76-99 IPB002494A 12.44 9.08e-10 25-58
1215	PR00858	Crustacean metallothionein signature II	PR00858B 5.93 1.48e-09 40-58 IPB000006 13.41 3.11e-09 36-81
1215	IPB001169	"Integrin beta, C-terminus"	IPB001169K 27.45 3.19e-09 42-84
1215	IPB002919	Trypsin Inhibitor-like cysteine rich domain	IPB002919A 15.56 3.57e-09 52-64 IPB002174A 30.51 4.15e-09 27-58 IPB001271 19.97 4.44e-09 58-86 IPB002494A 12.44 4.97e-09 32-65 PR01228C 5.69 5.03e-09 18-33 PR01228C 5.69 5.03e-09 22-37 IPB002174A 30.51 5.28e-09 19-50
1215	IPB000254	"Cellulose-binding domain, fungal type"	IPB000254 18.11 5.36e-09 28-58 IPB000006 13.41 5.59e-09 42-87 IPB002174A 30.51 5.72e-09 36-67 PR01228C 5.69 5.76e-09 27-42
1215	IPB000564	2Fe-2S Ferredoxin	IPB000564A 17.31 6.49e-09 1-19
1215	IPB000867	Insulin-like growth factor-binding protein	IPB000867B 11.44 6.55e-09 5-21 IPB002174A 30.51 6.62e-09 7-38
1215	IPB002867	Cysteine-rich domain (C6HC)	IPB002867D 24.88 7.19e-09 38-69 IPB000006 13.41 7.24e-09 50-95
1215	IPB000967	Zinc finger NF-X1 type	IPB000967D 10.42 7.37e-09 60-95 IPB001169K 27.45 7.81e-09 35-77 IPB000006 13.41 8.07e-09 3-48 IPB000006 13.41 8.07e-09 40-85 IPB002494A 12.44 8.35e-09 29-62 IPB000006 13.41 8.44e-09 55-100
1215	PR01117	CLC-6 chloride channel signature I	PR01117A 7.79 9.47e-09 51-63 IPB001271 19.97 9.51e-09 67-95 IPB002174A 30.51 9.77e-09 39-70
1215	IPB002221	WAP-type (Whey Acidic Protein) four-disulfide core domain	IPB002221B 17.12 1.00e-08 48-69
1218	PR00946	Mercury scavenger protein signature I	PR00946A 4.14 8.16e-09 6-24
1221	IPB002038	Osteopontin	IPB002038C 22.35 1.00e-40 119-160
1221	PR00216	Osteopontin signature I	PR00216A 11.45 9.71e-34 2-31 IPB002038B 15.58 2.06e-32 23-67

TABLE 3B

			PR00216C 9.12 5.85e-32 41-66 IPB002038A 12.23 5.15e-31 1-30 PR00216G 12.73 8.50e-30 231-256 PR00216F 12.92 1.62e-22 152-170 PR00216D 3.16 3.30e-18 88-102 PR00216E 6.95 3.81e-18 120-134 IPB002038D 9.52 5.50e-17 248-263 PR00216D 3.16 3.69e-12 82-96
1221	IPB003403	Herpesvirus immediate early protein	IPB003403E 17.25 9.26e-09 63-90
1222	IPB000215	Serpins	IPB000215A 13.01 9.14e-18 107-130 IPB000215D 15.35 3.74e-17 332-358 IPB000215E 15.36 6.68e-16 419-443 IPB000215C 13.90 7.88e-15 229-243
1223	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 3.52e-10 279-316 IPB003006A 17.51 7.75e-09 141-163
1225	IPB001241	DNA topoisomerase II family	IPB001241F 23.94 8.36e-37 399-447
1225	PR01158	Topoisomerase II signature VIII	PR01158H 13.39 5.50e-30 728-750 IPB001241G 14.13 1.00e-29 471-497 PR01158K 14.14 5.24e-27 947-973 PR01158G 9.37 5.91e-27 681-704
1225	IPB002205	"DNA gyrase/topoisomerase IV, subunit A"	IPB002205B 14.49 4.79e-24 684-719 IPB001241E 20.94 3.00e-22 295-321 PR01158I 13.95 7.00e-22 758-778 PR01158D 11.94 5.24e-21 489-504
1225	PR00418	DNA topoisomerase II family signature VI	PR00418F 13.13 3.40e-20 470-486 IPB001241B 10.04 2.71e-19 96-114 PR00418G 12.91 8.94e-19 488-505 IPB001241H 17.27 1.96e-18 732-755
1225	PR00615	CCAAT-binding transcription factor subunit A signature I	PR00615A 17.09 2.93e-18 243-261 PR01158J 13.56 3.45e-18 863-877 IPB002205D 10.13 3.54e-18 791-812 PR00615B 18.03 3.77e-18 631-649 PR00418C 9.38 1.82e-17 100-114 PR00418I 17.21 4.60e-17 550-566 IPB002205A 8.13 9.54e-17 653-671 PR00418A 13.58 7.65e-16 20-35 PR01158C 11.35 1.00e-15 443-456 PR01158E 8.11 2.29e-15 509-520 PR01158F 10.39 4.71e-15 556-568 PR00615C 17.93 8.50e-15 1072-1090 PR00418E 14.82 1.37e-14 397-411 IPB001241D 14.87 1.43e-14 252-265 PR00418B 12.37 2.57e-14 57-70 PR00418D 14.25 2.71e-14 252-265 PR01158A 7.61 4.60e-13 380-390 IPB002205C 11.89 5.09e-12 736-750 PR00418H 10.58 5.91e-12 508-520 IPB001241C 13.37 1.31e-11 154-166
1225	IPB000509	Ribosomal protein L36E	IPB000509B 20.29 7.85e-11 1140-1194 PR01158B 8.30 1.27e-10 395-402
1225	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 5.64e-09 1286-1310 IPB000135D 2.13 7.45e-09 1287-1311 IPB000135D 2.13 8.09e-09 1288-1312
1225	PR01469	Bacterial carbamate kinase signature V	PR01469E 10.60 8.43e-09 52-70 IPB000135D 2.13 8.73e-09 1284-1308
1226	IPB000873	AMP-dependent synthetase and ligase	IPB000873A 11.08 1.50e-12 248-263
1226	PR00154	AMP-binding signature I	PR00154A 8.79 5.14e-09 241-252

TABLE 3B

1227	IPB001043	"Vinculin, type 1"	IPB001043E 22.70 9.08e-09 136-173
1228	IPB001073	Complement C1q protein	IPB001073B 20.88 3.48e-24 96-130 IPB001073C 13.07 4.50e-13 163-182 IPB001073A 22.14 6.55e-13 42-76
1228	PR00007	Complement C1Q domain signature II	PR00007B 15.63 9.56e-13 116-135 IPB001073D 7.60 1.00e-11 195-204 PR00007D 9.66 2.00e-11 193-203 PR00007C 16.13 7.38e-11 163-184 PR00007A 20.64 3.04e-10 89-115
1230	IPB000906	ZU5 domain	IPB000906A 22.49 1.99e-15 274-316
1230	PR01415	Ankyrin repeat signature I	PR01415A 12.73 3.70e-15 381-393 IPB000906G 25.85 6.04e-12 900-948 IPB000906A 22.49 2.24e-11 893-935 PR01415A 12.73 1.00e-10 281-293 IPB000906F 35.93 1.61e-10 225-278 PR01415A 12.73 2.45e-10 796-808 IPB000906D 23.89 3.88e-10 3
1230	PR00665	Oxytocin receptor signature V	PR00665E 6.24 6.76e-09 756-769 IPB000906E 22.11 7.22e-09 278-318 PR01415B 10.23 7.75e-09 260-272 PR01415B 10.23 9.25e-09 227-239
1231	IPB001124	Lipid-binding serum glycoprotein	IPB001124C 25.71 7.71e-17 210-253 IPB001124D 21.85 5.71e-14 274-310
1232	IPB001124	Lipid-binding serum glycoprotein	IPB001124C 25.71 7.71e-17 210-253 IPB001124D 21.85 5.71e-14 274-310
1233	IPB001124	Lipid-binding serum glycoprotein	IPB001124C 25.71 7.71e-17 210-253 IPB001124D 21.85 5.71e-14 274-310
1234	PR00053	Fork head domain signature II	PR00053B 12.24 8.50e-09 523-540
1236	IPB000258	Bacterial ice-nucleation proteins octamer repeat	IPB000258G 8.61 7.77e-09 92-145
1237	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.57e-13 253-290
1240	IPB001627	Sema domain	IPB001627F 22.05 5.09e-29 255-288 IPB001627G 21.49 2.17e-28 311-344 IPB001627C 21.13 1.22e-21 162-193 IPB001627B 18.84 1.79e-21 117-145
1240	IPB002165	Plexin repeat	IPB002165C 18.49 3.45e-19 255-287 IPB001627I 10.67 6.57e-15 386-399 IPB001627A 16.97 5.26e-14 98-113 IPB001627H 10.22 1.35e-13 358-370 IPB001627K 13.76 7.92e-13 524-536 IPB001627J 11.43 1.22e-12 436-452 IPB002165C 18.49 3.64e-12 254-286 IPB002165D 14.72 3.65e-12 524-536 IPB001627D 16.04 6.70e-12 209-224 IPB002165B 13.59 7.57e-12 136-145 IPB001627E 8.70 9.59e-12 230-239
1247	PR00011	Type III EGF-like signature IV	PR00011D 12.12 8.93e-16 767-785 PR00011D 12.12 1.00e-15 550-568 PR00011B 13.08 5.06e-15 767-785 PR00011B 13.08 6.65e-15 289-307 PR00011D 12.12 6.67e-15 289-307 PR00011A 14.05 2.53e-14 289-307 PR00011D 12.12 5.86e-14 638-656 PR00011B 13.08 8.50e-14 550-568 PR00011B 13.08 1.93e-13 160-178 PR00011B 13.08 2.55e-13 203-221

421

TABLE 3B

			PR00011B 13.08 2.86e-13 421-439 PR00011D 12.12 3.83e-13 378-396 PR00011D 12.12 6.00e-13 421-439 PR00011A 14.05 7.83e-13 378-396 PR00011A 14.05 9.53e-13 203-221 PR00011B 13.08 9.53e-13 378-396 PR00011D 12.12 1.00e-12 810-828 PR00011B 13.08 1.59e-12 810-828 PR00011A 14.05 2.05e-12 550-568 PR00011D 12.12 3.02e-12 203-221 PR00011B 13.08 4.84e-12 638-656 PR00011D 12.12 5.50e-12 160-178 PR00011D 12.12 7.67e-12 507-525
1247	IPB000561	EGF-like domain	IPB000561 4.89 7.75e-12 210-218 PR00011D 12.12 8.29e-12 332-350 PR00011A 14.05 8.65e-12 421-439 PR00011A 14.05 1.55e-11 767-785 PR00011D 12.12 1.73e-11 593-611 PR00011A 14.05 3.08e-11 638-656 PR00011B 13.08 5.43e-11 593-611 PR00011D 12.12 6.66e-11 464-482 PR00011B 13.08 7.78e-11 332-350 PR00011D 12.12 7.82e-11 724-742
1247	IPB000034	Laminin B	IPB000034C 12.97 8.04e-11 210-228 PR00011A 14.05 8.34e-11 724-742 PR00011A 14.05 8.62e-11 160-178 PR00011B 13.08 9.03e-11 246-264 PR00011A 14.05 1.40e-10 810-828 PR00011B 13.08 1.53e-10 724-742 PR00011A 14.05 1.93e-10 507-525 PR00011D 12.12 2.25e-10 246-264 PR00011B 13.08 2.59e-10 507-525 PR00011A 14.05 4.04e-10 464-482
1247	IPB001774	Delta serrate ligand	IPB001774C 18.25 4.35e-10 115-157 IPB000561 4.89 4.75e-10 296-304 PR00011A 14.05 5.63e-10 246-264
1247	IPB001886	Laminin N-terminal (Domain VI)	IPB001886E 10.90 7.17e-10 294-310 PR00011D 12.12 8.20e-10 681-699 PR00011B 13.08 1.25e-09 464-482 IPB000561 4.89 1.64e-09 731-739 PR00011A 14.05 2.00e-09 332-350 PR00011A 14.05 2.75e-09 681-699
1247	PR00764	Complement C9 signature VI	PR00764F 15.74 3.96e-09 237-257
1247	IPB002174	Furin-like cysteine rich region	IPB002174A 30.51 4.60e-09 785-816 PR00011A 14.05 4.87e-09 593-611
1247	IPB002899	EB module	IPB002899A 6.67 6.32e-09 415-421 IPB002899A 6.67 6.32e-09 761-767
1247	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494A 12.44 6.32e-09 652-685
1247	IPB003884	Factor I membrane attack complex	IPB003884F 16.26 7.27e-09 587-602 IPB000034C 12.97 7.55e-09 296-314 IPB001886E 10.90 7.83e-09 772-788 IPB000561 4.89 8.71e-09 645-653 IPB000561 4.89 8.71e-09 688-696 PR00011B 13.08 8.77e-09 681-699 IPB000561 4.89 1.00e-08 253-261
1249	IPB002867	Cysteine-rich domain (C6HC)	IPB002867D 24.88 5.04e-18 129-160
1249	PR01475	Parkin signature IX	PR01475I 10.01 8.01e-09 86-108
1254	IPB002209	HBGF (heparin binding growth	IPB002209B 26.84 8.50e-31 90-128

422

TABLE 3B

		factor)/FGF (fibroblast growth factor) family	IPB002209C 23.35 1.00e-19 137-164
1254	PR00262	IL1/HBGF family signature I	PR00262A 25.25 4.38e-11 77-104
1254	PR00263	Heparin binding growth factor family signature IV	PR00263D 13.56 5.57e-11 106-125 PR00263C 8.53 7.51e-10 90-102 PR00262B 23.59 1.00e-08 108-128
1258	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 7.48e-10 165-202
1260	IPB000956	Stathmin	IPB000956B 9.49 7.36e-11 208-241
1260	PR00345	Stathmin family signature II	PR00345B 6.89 9.15e-11 207-235
1260	IPB000533	Tropomyosin	IPB000533C 10.81 3.06e-09 113-154
1261	IPB000215	Serpins	IPB000215D 15.35 5.03e-14 324-350 IPB000215A 13.01 2.91e-12 49-72 IPB000215C 13.90 5.00e-09 216-230
1262	PR01377	Claudin-1 signature I	PR01377A 7.94 1.00e-16 22-33
1263	PR00328	GTP-binding SAR1 protein signature I	PR00328A 12.43 5.14e-12 27-50 PR00328B 7.64 2.38e-11 55-79
1263	IPB000251	ADP-ribosylation factors family	IPB000251A 23.98 9.70e-09 55-108
1264	IPB001919	"Cellulose-binding domain, bacterial type"	IPB001919B 14.22 2.97e-09 270-294
1265	PR00258	Speract receptor signature II	PR00258B 7.94 3.00e-16 493-504 PR00258C 9.05 3.70e-14 62-72 PR00258C 9.05 7.30e-14 508-518 PR00258A 13.56 4.34e-13 474-490 PR00258D 14.29 2.66e-12 93-107 PR00258D 14.29 4.55e-12 539-553 PR00258A 13.56 7.20e-11 133-149 PR00258D 14.29 4.53e-10 294-308 PR00258A 13.56 6.22e-10 229-245 PR00258C 9.05 4.83e-09 163-173 PR00258E 14.06 5.72e-09 215-227 PR00258E 14.06 7.20e-09 562-574
1266	PR00258	Speract receptor signature II	PR00258B 7.94 3.00e-16 493-504 PR00258C 9.05 3.70e-14 62-72 PR00258C 9.05 7.30e-14 508-518 PR00258A 13.56 4.34e-13 474-490 PR00258D 14.29 2.66e-12 93-107 PR00258D 14.29 4.55e-12 539-553 PR00258A 13.56 7.20e-11 133-149 PR00258D 14.29 4.53e-10 294-308 PR00258A 13.56 6.22e-10 229-245 PR00258C 9.05 4.83e-09 163-173 PR00258E 14.06 5.72e-09 215-227 PR00258E 14.06 7.20e-09 562-574
1270	PR01305	Invasion protein B family signature IV	PR01305D 7.82 6.19e-09 423-436
1273	IPB001245	Tyrosine kinase catalytic domain	IPB001245A 22.45 1.00e-27 207-247
1273	IPB003527	MAP kinase	IPB003527C 14.70 2.94e-27 199-247
1273	IPB000961	Protein kinase C-terminal domain	IPB000961C 15.48 5.95e-22 214-248 IPB003527D 21.53 2.80e-17 256-297
1273	IPB001772	Kinase associated domain 1	IPB001772C 20.66 3.29e-17 202-232
1273	IPB000095	PAK-box /P21-Rho-binding	IPB000095E 17.62 6.35e-17 215-260
1273	IPB000861	PKN/rhopilin/rhotekin rho-binding repeat	IPB000861F 16.50 9.81e-16 208-262
1273	IPB000959	POLO box duplicated region	IPB000959B 15.68 3.01e-14 191-231
1273	IPB000494	"Epidermal growth-factor receptor (EGFR), L domain"	IPB000494C 24.40 7.88e-14 201-247 IPB001245B 21.68 6.19e-13 263-301

TABLE 3B

			IPB003527G 17.26 3.20e-10 360-397 IPB000961D 21.23 5.27e-10 259-300 IPB000961A 16.82 3.33e-09 102-136
1273	PR00109	Tyrosine kinase catalytic domain signature II	PR00109B 11.07 7.75e-09 214-232
1275	IPB001762	Disintegrin	IPB001762A 23.93 4.33e-23 458-498
1275	IPB002870	Reprolysin family propeptide	IPB002870B 24.73 3.54e-20 131-169
1275	PR00289	Disintegrin signature I	PR00289A 14.29 1.16e-14 474-493 IPB002870F 18.81 3.03e-14 402-426 IPB002870E 11.90 2.46e-12 361-373 IPB001762B 10.06 3.40e-12 505-515 IPB001762A 23.93 9.20e-11 426-466
1275	IPB000130	"Neutral zinc metallopeptidases, zinc-binding region"	IPB000130 5.86 1.56e-10 359-369
1275	PR00138	Matrixin signature IV	PR00138D 14.57 2.54e-10 359-384 IPB002870D 16.31 4.77e-10 327-342
1275	IPB001774	Delta serrate ligand	IPB001774C 18.25 5.31e-10 677-719
1275	PR00480	Astacin family signature II	PR00480B 14.35 5.57e-10 354-372
1275	PR00436	Interleukin-8 signature I	PR00436A 15.20 7.43e-10 5-28
1275	IPB001818	Matrixin	IPB001818D 14.91 1.72e-09 353-384 PR00289B 11.74 3.80e-09 503-515 IPB002870A 12.22 6.54e-09 85-101
1275	IPB003306	WIF domain	IPB003306E 25.51 7.40e-09 654-699
1275	PR01236	Tumour necrosis factor beta (lymphotoxin-alpha) signature I	PR01236A 4.92 7.49e-09 17-33 IPB002870C 11.01 9.64e-09 295-305
1277	PR01415	Ankyrin repeat signature I	PR01415A 12.73 1.00e-12 341-353 PR01415A 12.73 2.29e-11 302-314
1277	PR01256	Otx1 transcription factor signature II	PR01256B 5.92 4.44e-09 431-443 PR01256B 5.92 9.39e-09 432-444
1278	PR00756	Membrane alanyl dipeptidase (M1) family signature IV	PR00756D 10.78 7.75e-18 412-427 PR00756A 12.71 1.45e-17 245-260 PR00756B 15.53 2.04e-14 297-312 PR00756E 10.37 5.68e-09 431-443
1278	IPB000130	"Neutral zinc metallopeptidases, zinc-binding region"	IPB000130 5.86 6.57e-09 412-422
1278	IPB002594	Glycoside hydrolase family 12	IPB002594A 4.24 1.00e-08 26-35
1288	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 6.85e-13 252-266
1288	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 5.64e-09 164-177
1288	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.19e-09 348-385 PR00019B 11.42 8.91e-09 112-125
1290	PR00019	Leucine-rich repeat signature II	PR00019B 11.42 4.18e-12 83-96 PR00019A 11.72 1.00e-10 86-99 PR00019A 11.72 1.67e-10 111-124
1290	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 7.43e-10 131-145
1290	IPB000267	Asparaginase/glutaminase family	IPB000267A 12.78 7.67e-09 11-27
1290	PR01528	EDG-4 lysophosphatidic acid receptor signature II	PR01528B 3.89 8.48e-09 130-144
1292	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 5.85e-09 195-232
1293	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 5.85e-09 195-232
1295	IPB001073	Complement C1q protein	IPB001073B 20.88 6.35e-20 92-126
1295	PR00007	Complement C1Q domain signature III	PR00007C 16.13 5.93e-14 159-180 PR00007B 15.63 1.66e-13 112-131 IPB001073C 13.07 2.25e-13 159-178

424

TABLE 3B

			IPB001073D 7.60 6.40e-12 191-200 IPB001073A 22.14 4.67e-11 32-66 PR00007D 9.66 6.29e-10 189-199 PR00007A 20.64 3.68e-09 86-112
1295	PR00513	5-hydroxytryptamine 1B receptor signature IV	PR00513D 10.60 9.80e-09 50-67
1296	PR01481	Neurotensin type 2 receptor signature VI	PR01481F 11.66 8.46e-28 236-259 PR01481E 6.05 7.87e-25 214-235 PR01481C 15.05 1.00e-17 150-163
1296	PR01479	Neurotensin receptor signature II	PR01479B 12.40 2.43e-17 89-101 PR01481A 7.58 3.54e-16 1-13 PR01481B 6.68 1.45e-15 14-26 PR01481D 4.62 2.19e-15 164-175 PR01479E 8.74 3.70e-15 305-315 PR01479D 13.10 6.57e-14 294-304 PR01479A 8.89 1.00e-13 29-39
1296	PR00237	Rhodopsin-like GPCR superfamily signature VI	PR00237F 14.34 9.33e-13 269-293 PR00237G 19.23 4.44e-12 314-340
1296	PR00665	Oxytocin receptor signature IV	PR00665D 10.30 1.32e-11 108-124 PR01479F 8.03 5.19e-11 342-352 PR00237A 9.81 7.33e-10 34-58 PR00237D 9.76 7.43e-10 125-146
1297	IPB001101	Plectin repeat	IPB001101C 6.05 3.42e-35 894-946
1297	IPB001589	Actinin-type actin-binding domain	IPB001589C 16.73 1.78e-31 285-316 IPB001589D 26.07 2.55e-27 340-383 IPB001101M 9.29 7.80e-27 1607-1657 IPB001101Z 7.76 2.12e-25 3013-3066 IPB001101B 12.20 1.00e-24 791-844 IPB001101F 10.86 3.20e-22 1078-1126 IPB001101E 6.00 7
1297	IPB002017	Spectrin repeat	IPB002017A 14.19 3.25e-11 246-262 IPB001101Q 7.28 8.69e-11 2855-2892 IPB001101S 8.38 9.52e-11 2695-2738 IPB001101N 4.86 2.32e-10 1779-1833 IPB001101N 4.86 3.81e-10 1758-1812 IPB001101N 4.86 3.87e-10 1737-1791 IPB001101R 5.90 3.91e-10 3112-3165 IPB001101T 7.36 5.01e-10 2720-2774 IPB001101W 10.36 5.46e-10 3033-3062 IPB001101T 7.36 5.53e-10 3067-3121 IPB001101R 5.90 2.07e-09 2727-2780
1297	IPB000237	GRIP domain	IPB000237B 30.66 2.76e-09 2392-2442 IPB001101Q 7.28 3.27e-09 3166-3203
1297	IPB001664	Intermediate filament proteins	IPB001664B 17.44 5.92e-09 1742-1781 IPB001101O 8.21 6.25e-09 1767-1800
1297	IPB002079	"Gag polyprotein, inner coat protein p12"	IPB002079J 10.53 6.85e-09 1766-1794
1297	IPB001715	Calponin homology (CH) domain	IPB001715A 10.74 7.00e-09 241-251 IPB001101W 10.36 7.63e-09 2798-2827 IPB001589E 11.55 8.94e-09 389-398
1297	IPB003865	Prolyl 4-hydroxylase alpha subunit C-terminus	IPB003865A 20.35 9.33e-09 2093-2137 IPB001101X 9.00 9.86e-09 3063-3096
1298	IPB001101	Plectin repeat	IPB001101C 6.05 3.42e-35 906-958
1298	IPB001589	Actinin-type actin-binding domain	IPB001589C 16.73 1.78e-31 297-328 IPB001589D 26.07 2.55e-27 352-395 IPB001101M 9.29 7.80e-27 1619-1669 IPB001101Z 7.76 2.12e-25 3025-3078

TABLE 3B

			IPB001101B 12.20 1.00e-24 803-856 IPB001101F 10.86 3.20e-22 1090-1138 IPB001101E 6.00 7
1298	IPB002017	Spectrin repeat	IPB002017A 14.19 3.25e-11 246-262 IPB001101Q 7.28 8.69e-11 2867-2904 IPB001101S 8.38 9.52e-11 2707-2750 IPB001101N 4.86 2.32e-10 1791-1845 IPB001101N 4.86 3.81e-10 1770-1824 IPB001101N 4.86 3.87e-10 1749-1803 IPB001101R 5.90 3.91e-10 3124-3177 IPB001101T 7.36 5.01e-10 2732-2786 IPB001101W 10.36 5.46e-10 3045-3074 IPB001101T 7.36 5.53e-10 3079-3133 IPB001101R 5.90 2.07e-09 2739-2792
1298	IPB000237	GRIP domain	IPB000237B 30.66 2.76e-09 2404-2454 IPB001101Q 7.28 3.27e-09 3178-3215
1298	IPB001664	Intermediate filament proteins	IPB001664B 17.44 5.92e-09 1754-1793 IPB001101O 8.21 6.25e-09 1779-1812
1298	IPB002079	"Gag polyprotein, inner coat protein p12"	IPB002079J 10.53 6.85e-09 1778-1806
1298	IPB001715	Calponin homology (CH) domain	IPB001715A 10.74 7.00e-09 241-251 IPB001101W 10.36 7.63e-09 2810-2839 IPB001589E 11.55 8.94e-09 401-410
1298	IPB003865	Prolyl 4-hydroxylase alpha subunit C-terminus	IPB003865A 20.35 9.33e-09 2105-2149 IPB001101X 9.00 9.86e-09 3075-3108
1306	IPB000998	MAM domain	IPB000998C 18.63 9.65e-15 510-525 IPB000998D 18.66 2.41e-14 575-598 IPB000998B 17.20 4.55e-10 430-442
1306	PR00020	MAM domain signature I	PR00020A 20.48 7.62e-10 428-446 PR00020C 12.01 4.78e-09 509-520
1308	IPB001552	Acyl-CoA dehydrogenase	IPB001552E 22.77 2.46e-19 726-766 IPB001552D 24.88 5.35e-19 635-677 IPB001552C 25.04 7.75e-15 581-621 IPB001552B 18.05 3.19e-12 530-552 IPB001552A 11.25 6.90e-10 503-514
1309	IPB001552	Acyl-CoA dehydrogenase	IPB001552E 22.77 2.46e-19 708-748 IPB001552D 24.88 5.35e-19 617-659 IPB001552C 25.04 7.75e-15 563-603 IPB001552B 18.05 3.19e-12 512-534 IPB001552A 11.25 6.90e-10 485-496
1310	IPB002524	Cation efflux family	IPB002524B 23.89 5.20e-17 86-125
1310	IPB003452	Stem cell factor	IPB003452B 19.11 6.63e-09 145-193
1311	PR00215	Neuromodulin signature III	PR00215C 13.82 7.58e-10 743-763
1311	PR00194	Tropomyosin signature IV	PR00194D 9.54 7.19e-09 622-645
1311	IPB001422	Neuromodulin (GAP-43)	IPB001422A 13.23 7.43e-09 718-762
1314	IPB000569	HECT domain (Ubiquitin-protein ligase)	IPB000569C 20.19 8.94e-30 2270-2299
1314	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 9.00e-17 361-385 IPB000135D 2.13 7.04e-16 370-394 IPB000135D 2.13 3.70e-15 360-384 IPB000135D 2.13 5.50e-15 364-388 IPB000135D 2.13 7.43e-15 367-391 IPB000135D 2.13 7.94e-15 365-389 IPB000569A 16.82 8.58e-15 2
1314	IPB001580	Calreticulin family	IPB001580F 2.93 5.50e-10 370-379
1314	IPB001990	Granins (chromogranin or secretogranin)	IPB001990C 33.59 6.26e-10 352-399 IPB001580F 2.93 7.75e-10 369-378

426
TABLE 3B

			IPB000135D 2.13 8.34e-10 351-375 IPB000569B 18.58 8.92e-10 2233-2249
1314	IPB003403	Herpesvirus immediate early protein	IPB003403E 17.25 8.97e-10 359-386
1314	IPB002889	WSC domain	IPB002889B 11.76 2.88e-09 1392-1438 IPB000135D 2.13 4.09e-09 381-405 IPB000135D 2.13 4.18e-09 352-376 IPB000135D 2.13 4.36e-09 353-377 IPB002889B 11.76 4.66e-09 1440-1486
1314	IPB002000	Lysosome-associated membrane glycoprotein (Lamp)	IPB002000D 5.87 6.26e-09 1429-1442 IPB000135D 2.13 6.27e-09 349-373 IPB001580F 2.93 6.40e-09 374-383 IPB000135D 2.13 6.45e-09 382-406 IPB002889B 11.76 6.81e-09 1458-1504 IPB002000D 5.87 7.11e-09 1434-1447 IPB002889B 11.76 7.47e-09 1417-1463 IPB001990C 33.59 7.51e-09 347-394 IPB000135D 2.13 8.36e-09 350-374 IPB002889B 11.76 9.53e-09 1402-1448
1314	IPB000637	HMG-I and HMG-Y DNA-binding domain (A+T-hook)	IPB000637B 14.21 9.73e-09 369-387
1314	PR01073	Presenilin 1 signature III	PR01073C 1.45 9.89e-09 367-378
1317	PR01145	Thyrotropin receptor precursor signature I	PR01145A 6.74 9.10e-11 3-22
1317	PR01472	Intercellular adhesion molecule/vascular cell adhesion molecule-1 signature I	PR01472A 16.78 7.66e-09 35-51
1321	PR00019	Leucine-rich repeat signature II	PR00019B 11.42 7.88e-12 335-348 PR00019B 11.42 1.33e-10 477-490 PR00019A 11.72 4.00e-10 480-493 PR00019A 11.72 4.33e-10 338-351
1321	IPB001580	Calreticulin family	IPB001580F 2.93 4.94e-10 648-657 IPB001580F 2.93 4.94e-10 649-658 IPB001580F 2.93 4.94e-10 650-659 PR00019B 11.42 5.33e-10 167-180 PR00019A 11.72 4.00e-09 454-467
1321	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 4.64e-09 637-661 PR00019B 11.42 7.55e-09 193-206 PR00019B 11.42 7.55e-09 309-322 PR00019B 11.42 7.82e-09 451-464 IPB000135D 2.13 8.55e-09 635-659
1322	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 9.14e-12 297-334
1322	IPB001000	Glycoside hydrolase family 10	IPB001000H 10.38 7.80e-09 8-21
1323	IPB001000	Glycoside hydrolase family 10	IPB001000H 10.38 7.80e-09 8-21
1324	IPB003884	Factor I membrane attack complex	IPB003884A 12.20 7.06e-09 34-45
1328	PR00258	Speract receptor signature II	PR00258B 7.94 5.00e-16 654-665 PR00258B 7.94 6.50e-16 30-41 PR00258B 7.94 6.50e-16 204-215 PR00258A 13.56 9.70e-14 635-651 PR00258B 7.94 2.58e-13 316-327 PR00258E 14.06 4.16e-13 491-503 PR00258A 13.56 5.63e-13 402-418 PR00258A 13.56 6.14e-13 185-201 PR00258B 7.94 6.62e-13 421-432 PR00258C 9.05 9.18e-13 45-55 PR00258A 13.56 1.22e-12 11-27 PR00258A 13.56 1.22e-12 297-313

427

TABLE 3B

			PR00258E 14.06 1.98e-12 99-111 PR00258E 14.06 9.22e-12 273-285 PR00258D 14.29 2.00e-11 468-482 PR00258D 14.29 3.20e-11 700-714 PR00258D 14.29 2.76e-10 250-264 PR00258C 9.05 4.95e-10 219-229 PR00258C 9.05 4.95e-10 331-341 PR00258E 14.06 5.42e-10 385-397 PR00258D 14.29 8.06e-10 362-376 PR00258C 9.05 7.51e-09 436-446
1333	IPB000970	"Developmental signaling protein, Wnt-1 family"	IPB000970E 22.74 1.00e-40 202-255 IPB000970F 23.43 1.51e-40 307-355 IPB000970C 13.22 2.80e-25 101-132 IPB000970B 14.73 6.14e-23 65-88
1333	PR01349	Wnt protein signature IV	PR01349D 8.90 3.81e-20 222-237 IPB000970D 13.85 3.48e-17 167-186 PR01349C 10.34 3.86e-15 167-179 PR01349A 11.18 8.55e-14 103-117 PR01349B 10.00 3.32e-12 122-135 PR01349E 12.39 5.61e-11 283-294
1333	IPB001073	Complement C1q protein	IPB001073A 22.14 4.20e-10 137-171 IPB000970A 13.08 5.78e-10 41-56
1335	PR00245	Olfactory receptor signature I	PR00245A 10.98 8.92e-11 59-70
1335	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 3.61e-09 18-30
1337	IPB001522	"Fatty acid desaturase, type 1"	IPB001522D 12.81 1.00e-40 119-154 IPB001522F 22.32 1.00e-40 241-295 IPB001522E 20.55 5.85e-36 163-216 IPB001522C 14.10 2.89e-33 81-117
1337	PR00075	Fatty acid desaturase family 1 signature IV	PR00075D 13.27 3.57e-33 131-160 PR00075C 10.51 3.40e-22 94-114 PR00075G 10.50 6.62e-20 268-282 PR00075E 11.60 6.46e-18 192-210 PR00075A 16.73 9.44e-17 47-67 PR00075F 14.62 8.81e-16 225-246 PR00075B 13.44 4.56e-14 71-93 IPB001522B 29.55 6.82e-12 29-80
1339	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 2.57e-17 46-70 IPB000135D 2.13 9.86e-17 43-67 IPB000135D 2.13 6.10e-16 45-69 IPB000135D 2.13 1.77e-15 47-71 IPB000135D 2.13 2.93e-15 44-68 IPB000135D 2.13 3.83e-15 41-65 IPB000135D 2.13 2.95e-14 48-72 IPB000135D 2.13 7.93e-14 42-66 IPB000135D 2.13 7.81e-13 49-73
1339	IPB001422	Neuromodulin (GAP-43)	IPB001422C 16.82 3.41e-11 40-75 IPB000135D 2.13 9.08e-11 40-64 IPB000135D 2.13 9.69e-11 50-74
1339	IPB001580	Calreticulin family	IPB001580F 2.93 1.00e-10 50-59 IPB000135D 2.13 2.17e-10 51-75 IPB000135D 2.13 3.15e-10 39-63 IPB001580F 2.93 4.94e-10 57-66 IPB001580F 2.93 4.94e-10 58-67 IPB001580F 2.93 5.50e-10 56-65 IPB001580F 2.93 6.06e-10 54-63 IPB001580F 2.93 7.75e-10 49-58 IPB001422C 16.82 7.99e-10 43-78

428
TABLE 3B

			IPB001422C 16.82 8.58e-10 42-77 IPB000135D 2.13 8.63e-10 38-62 IPB001580F 2.93 8.88e-10 51-60 IPB001422C 16.82 9.05e-10 46-81 IPB001580F 2.93 9.44e-10 59-68 IPB001422C 16.82 5.61e-09 48-83 IPB000135D 2.13 6.27e-09 37-61 IPB001422C 16.82 6.40e-09 44-79 IPB001580F 2.93 6.40e-09 52-61 IPB001422C 16.82 8.99e-09 47-82
1339	IPB000637	HMG-I and HMG-Y DNA-binding domain (A+T-hook)	IPB000637B 14.21 1.00e-08 45-63 IPB001580F 2.93 1.00e-08 61-70
1340	IPB004000	Actin and actin-like	IPB004000C 8.66 4.86e-20 137-191 IPB004000D 13.38 5.70e-16 267-321
1340	PR00190	Actin signature VI	PR00190F 7.36 2.20e-14 135-154 IPB004000A 9.97 4.64e-13 5-43 IPB004000B 6.57 5.80e-12 83-133
1341	PR01333	Two pore domain K ⁺ channel signature I	PR01333A 18.74 4.00e-18 125-153
1341	PR01463	EAG/ELK/ERG potassium channel family signature VI	PR01463F 4.09 1.95e-12 243-260 PR01333B 10.39 9.71e-10 255-264
1341	PR01526	EDG-6 sphingosine 1-phosphate receptor signature IV	PR01526D 5.56 9.71e-09 1-16
1343	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.20e-10 348-385
1344	IPB000998	MAM domain	IPB000998C 18.63 1.95e-12 833-848 IPB000998B 17.20 1.62e-11 761-773
1344	PR00020	MAM domain signature I	PR00020A 20.48 3.62e-11 759-777 PR00020C 12.01 8.12e-10 832-843 IPB000998D 18.66 9.61e-10 898-921
1344	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006A 17.51 7.11e-09 354-376
1344	PR00096	Glutamine amidotransferase superfamily signature III	PR00096C 15.85 9.28e-09 534-547
1345	IPB002350	Kazal-type serine protease inhibitor family	IPB002350 31.78 3.92e-13 127-167
1345	IPB000867	Insulin-like growth factor-binding protein	IPB000867B 11.44 1.37e-12 75-91
1345	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 3.88e-10 231-268
1345	IPB002328	Zinc-containing alcohol dehydrogenase	IPB002328C 11.03 8.84e-10 76-90
1346	IPB000224	Vesiculovirus phosphoprotein	IPB000224A 7.26 6.74e-10 437-470
1346	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 7.16e-10 430-454
1346	PR00449	Transforming protein P21 ras signature I	PR00449A 12.48 8.16e-10 83-104
1346	PR00326	GTP1/OBG GTP-binding protein family signature I	PR00326A 8.70 9.13e-10 85-105 IPB000135D 2.13 3.09e-09 434-458
1346	IPB000619	Guanylate kinase	IPB000619A 18.08 4.21e-09 85-102
1346	PR00905	Hypothetical mycoplasma lipoprotein (MG045) signature VIII	PR00905H 6.88 5.89e-09 343-363
1346	PR00364	Disease resistance protein signature I	PR00364A 8.29 7.14e-09 84-99
1346	PR00094	Adenylate kinase signature I	PR00094A 9.62 9.57e-09 86-99
1346	PR00918	Calicivirus non-structural polyprotein family signature I	PR00918A 13.81 9.69e-09 79-99
1346	IPB000795	GTP-binding elongation factor	IPB000795A 10.67 9.77e-09 84-99

429

TABLE 3B

			IPB000135D 2.13 9.82e-09 429-453
1348	PR00406	Cytochrome B5 reductase signature VI	PR00406F 4.29 4.86e-11 158-166
1348	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.48e-11 473-510 IPB003006B 20.23 4.60e-10 291-328 IPB003006B 20.23 3.08e-09 190-227
1348	PR00014	Fibronectin type III repeat signature IV	PR00014D 15.12 8.83e-09 891-905
1349	PR00698	C.elegans Srg family integral membrane protein signature V	PR00698E 14.65 2.76e-09 97-122
1349	IPB002146	ATP synthase B/B' CF(0)	IPB002146 21.39 6.94e-09 174-212
1350	IPB000215	Serpins	IPB000215D 15.35 1.41e-18 311-337 IPB000215A 13.01 8.29e-18 72-95 IPB000215C 13.90 1.53e-15 207-221 IPB000215E 15.36 7.00e-13 378-402 IPB000215B 9.87 4.68e-11 180-192
1352	IPB001737	Ribosomal RNA adenine dimethylase	IPB001737A 27.11 8.54e-10 134-179
1355	IPB000906	ZU5 domain	IPB000906G 25.85 6.28e-10 164-212 IPB000906A 22.49 3.16e-09 58-100
1356	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 6.54e-13 385-423
1356	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 3.97e-11 389-443
1356	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 2.22e-10 381-422 IPB001245A 22.45 3.18e-10 332-372
1356	IPB001359	Synapsin	IPB001359H 22.58 7.12e-10 696-746 IPB001359H 22.58 4.84e-09 695-745
1356	IPB002889	WSC domain	IPB002889B 11.76 6.81e-09 1510-1556 IPB002889B 11.76 9.25e-09 1491-1537
1357	IPB001359	Synapsin	IPB001359H 22.58 7.12e-10 289-339 IPB001359H 22.58 4.84e-09 288-338
1357	IPB002889	WSC domain	IPB002889B 11.76 6.81e-09 1103-1149 IPB002889B 11.76 9.25e-09 1084-1130
1358	PR00237	Rhodopsin-like GPCR superfamily signature VII	PR00237G 19.23 9.64e-15 41-67
1358	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276D 9.40 5.05e-12 51-67 IPB000276C 8.03 8.50e-11 8-19
1359	PR01041	Methionyl-tRNA synthetase signature V	PR01041E 16.72 2.69e-17 306-321 PR01041D 11.02 7.43e-13 276-287 PR01041A 11.40 8.68e-13 47-60
1359	IPB001412	Aminoacyl-transfer RNA synthetases class-I	IPB001412B 6.33 8.71e-12 344-354 PR01041B 11.59 4.06e-09 82-96
1359	PR01038	Arginyl-tRNA synthetase signature II	PR01038B 9.12 7.68e-09 59-75
1360	IPB000353	"Class II histocompatibility antigen, beta chain, beta-1 domain"	IPB000353A 18.51 7.30e-27 42-91
1363	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.83e-11 374-411
1365	PR01360	Interleukin-1 receptor antagonist precursor IL-1RA signature VI	PR01360F 14.44 9.86e-18 116-134
1365	PR00264	Interleukin-1 precursor family signature III	PR00264C 19.37 4.90e-16 108-123 PR01360E 9.69 9.33e-13 95-115
1365	IPB000975	Interleukin-1	IPB000975E 28.12 3.57e-12 95-134
1365	PR01357	Interleukin-1 alpha/beta precursor family signature VI	PR01357F 17.87 7.15e-10 108-123 PR00264A 18.63 9.85e-09 55-75
1366	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 1224-1251 IPB001599F 18.95 7.00e-24 786-815 IPB001599H 18.42 6.40e-20 999-1026 IPB001599N 24.85 7.69e-20 1417-1449

430
TABLE 3B

			IPB001599A 10.97 9.69e-18 123-141
1366	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 1237-1251 IPB001599M 13.29 4.71e-13 1364-1375 IPB001599G 13.87 8.94e-13 967-976 IPB001599B 7.45 4.89e-12 209-221 IPB001599D 11.61 6.90e-12 729-739 IPB001599J 20.99 3.00e-11 1065-1090 IPB001599I 10.83 7.60e-11 1034-1043 IPB001599K 8.15 1.46e-10 1194-1205 IPB001599C 14.40 3.55e-09 236-252 IPB001599E 11.06 9.77e-09 756-765
1368	IPB001526	Ly-6/u-PAR domain	IPB001526C 13.04 7.55e-15 90-105 IPB001526A 13.24 9.14e-11 12-27 IPB001526B 12.26 7.75e-10 46-55
1967	IPB001400	Somatotropin hormone family	IPB001400B 23.62 1.90e-28 99-135 IPB001400A 14.85 4.91e-16 55-78
1967	PR00836	Somatotropin hormone family signature II	PR00836B 17.50 2.44e-14 121-139 PR00836A 15.53 2.35e-13 99-112
1968	IPB001400	Somatotropin hormone family	IPB001400B 23.62 1.90e-28 99-135 IPB001400A 14.85 4.91e-16 55-78
1968	PR00836	Somatotropin hormone family signature II	PR00836B 17.50 2.44e-14 121-139 PR00836A 15.53 2.35e-13 99-112
1969	IPB001400	Somatotropin hormone family	IPB001400B 23.62 1.90e-28 99-135 IPB001400A 14.85 4.91e-16 55-78
1969	PR00836	Somatotropin hormone family signature II	PR00836B 17.50 2.44e-14 121-139 PR00836A 15.53 2.35e-13 99-112
1970	IPB001400	Somatotropin hormone family	IPB001400B 23.62 1.90e-28 99-135 IPB001400A 14.85 4.91e-16 55-78
1970	PR00836	Somatotropin hormone family signature II	PR00836B 17.50 2.44e-14 121-139 PR00836A 15.53 2.35e-13 99-112
1971	IPB000215	Serpins	IPB000215E 15.36 5.76e-17 425-449 IPB000215A 13.01 3.42e-15 111-134 IPB000215D 15.35 8.05e-11 346-372 IPB000215C 13.90 1.29e-10 241-255 IPB000215B 9.87 6.04e-10 214-226
1972	PR00390	Phospholipase C signature I	PR00390A 14.24 6.34e-20 2-20
1973	IPB000734	Lipase	IPB000734 10.25 8.50e-09 468-482
1977	IPB000689	UbiH/COQ6 monooxygenase family	IPB000689D 28.07 7.83e-39 377-427 IPB000689B 27.03 9.59e-28 217-251 IPB000689C 18.76 3.74e-24 262-286 IPB000689A 9.11 1.25e-11 52-64
1977	PR00420	Aromatic-ring hydroxylase (flavoprotein monooxygenase) signature III	PR00420C 12.44 8.53e-11 373-388
1977	PR01001	FAD-dependent glycerol-3-phosphate dehydrogenase family signature I	PR01001A 8.45 1.60e-09 51-63 PR00420A 15.97 3.95e-09 52-74 PR00420B 13.97 8.53e-09 215-230
1980	IPB000345	Cytochrome c family heme-binding site	IPB000345 9.03 7.19e-09 153-165
1982	IPB002610	Rhomboid family	IPB002610C 5.81 3.81e-10 262-272 IPB002610B 5.33 6.81e-09 203-213
1984	IPB001124	Lipid-binding serum glycoprotein	IPB001124D 21.85 2.50e-12 251-287 IPB001124C 25.71 5.08e-11 184-227
1985	IPB000817	Prion protein	IPB000817A 8.34 6.40e-09 70-112 IPB000817A 8.34 8.67e-09 64-106
1988	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442F 15.05 1.00e-40 585-628 IPB001442C 14.98 4.82e-40 498-532

431

TABLE 3B

			IPB001442A 26.12 4.09e-39 259-311 IPB001442D 15.34 1.00e-34 533-564
1988	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 1.93e-27 300-353 IPB001442A 26.12 8.93e-27 103-155 IPB001442A 26.12 9.69e-27 106-158 IPB001442A 26.12 4.19e-26 368-420 IPB000885A 11.46 4.80e-26 363-400 IPB001442A 26.12 6.52e-26 112-164 IPB001442A 26.12 9.71
1988	IPB001073	Complement C1q protein	IPB001073A 22.14 9.18e-19 374-408 IPB000885B 19.15 9.40e-19 309-362 IPB000885B 19.15 9.40e-19 373-426 IPB001442A 26.12 9.42e-19 265-317 IPB001442A 26.12 9.77e-19 133-185 IPB000885B 19.15 1.12e-18 81-134 IPB001442A 26.12 1.33e
1988	IPB001285	Synaptophysin/synaptoporin	IPB001285F 6.39 4.08e-09 340-384 IPB000885B 19.15 4.11e-09 48-101 IPB000885B 19.15 4.35e-09 174-227 IPB001442B 12.38 4.41e-09 257-277 IPB001442B 12.38 4.41e-09 417-437 IPB000885B 19.15 4.68e-09 147-200 IPB000885B 19.15 4.68e-
1988	IPB000817	Prion protein	IPB000817A 8.34 7.73e-09 258-300 IPB001073A 22.14 7.75e-09 76-110 IPB001442B 12.38 7.81e-09 25-45 IPB001073A 22.14 7.89e-09 151-185 IPB001073A 22.14 8.31e-09 416-450 IPB000817A 8.34 8.39e-09 255-297 IPB001442B 12.38 8.42e-09 363-383 IPB001442A 26.12 8.59e-09 160-212 IPB001442A 26.12 8.90e-09 40-92 IPB001442B 12.38 8.91e-09 429-449 IPB000885B 19.15 8.94e-09 324-377 IPB001073A 22.14 9.30e-09 82-116 IPB001073A 22.14 9.30e-09 307-341 IPB001442B 12.38 9.64e-09 323-343 IPB001073A 22.14 9.72e-09 148-182 IPB000885B 19.15 9.84e-09 412-465
1989	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033D 30.18 1.18e-14 111-149 IPB000033D 30.18 6.25e-11 67-105 IPB000033C 11.58 6.40e-10 135-149 IPB000033C 11.58 8.07e-09 48-62 IPB000033C 11.58 8.07e-09 91-105
1990	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033D 30.18 1.18e-14 111-149 IPB000033D 30.18 6.25e-11 67-105 IPB000033C 11.58 6.40e-10 135-149 IPB000033C 11.58 8.07e-09 48-62 IPB000033C 11.58 8.07e-09 91-105
1992	PR00205	Cadherin signature II	PR00205B 20.09 4.94e-14 114-143 PR00205D 12.22 9.31e-14 198-217 PR00205F 19.57 1.53e-12 167-193 PR00205D 12.22 8.20e-12 93-112 PR00205G 13.05 2.46e-11 201-218 PR00205G 13.05 3.93e-10 96-113
1992	IPB002126	Cadherin domain	IPB002126B 12.04 7.68e-10 102-119 PR00205A 17.38 8.15e-09 160-179
1993	PR00205	Cadherin signature II	PR00205B 20.09 4.94e-14 114-143

432

TABLE 3B

			PR00205D 12.22 9.31e-14 198-217 PR00205F 19.57 1.53e-12 167-193 PR00205D 12.22 8.20e-12 93-112 PR00205G 13.05 2.46e-11 201-218 PR00205G 13.05 3.93e-10 96-113
1993	IPB002126	Cadherin domain	IPB002126B 12.04 7.68e-10 102-119 PR00205A 17.38 8.15e-09 160-179
1994	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469J 8.97 3.52e-12 17-33
1995	PR01534	Vomerolnasal type 1 receptor family signature V	PR01534E 7.16 1.23e-09 5-19
1996	IPB000221	Protamine P1	IPB000221 5.48 2.97e-12 124-150 IPB000221 5.48 9.30e-12 113-139 IPB000221 5.48 2.19e-11 153-179 IPB000221 5.48 2.59e-11 114-140 IPB000221 5.48 3.91e-11 128-154
1996	IPB000492	Protamine 2 (PRM2)	IPB000492B 5.26 5.88e-11 148-182 IPB000221 5.48 6.16e-11 142-168 IPB000221 5.48 6.43e-11 149-175 IPB000221 5.48 7.62e-11 110-136 IPB000492B 5.26 9.35e-11 129-163 IPB000492B 5.26 9.35e-11 152-186 IPB000221 5.48 2.73e-10 168-194 IPB000221 5.48 4.70e-10 112-138 IPB000221 5.48 4.70e-10 144-170 IPB000492B 5.26 6.97e-10 153-187 IPB000492B 5.26 8.12e-10 156-190 IPB000492B 5.26 8.53e-10 155-189 IPB000221 5.48 8.89e-10 151-177 IPB000492B 5.26 9.06e-10 128-162 IPB000492B 5.26 9.69e-10 150-184 IPB000221 5.48 1.00e-09 133-159 IPB000221 5.48 1.46e-09 115-141 IPB000221 5.48 3.31e-09 159-185 IPB000221 5.48 3.31e-09 172-198 IPB000492B 5.26 3.84e-09 125-159 IPB000221 5.48 5.15e-09 157-183 IPB000221 5.48 5.27e-09 102-128
1996	PR00055	HIV TAT domain signature III	PR00055C 9.12 5.92e-09 66-82 IPB000221 5.48 6.19e-09 166-192 IPB000492B 5.26 6.38e-09 144-178 IPB000492B 5.26 6.67e-09 157-191 IPB000221 5.48 6.88e-09 147-173 IPB000221 5.48 6.88e-09 161-187 IPB000492B 5.26 7.75e-09 127-161 IPB000492B 5.26 8.34e-09 115-149
1996	IPB000271	Ribosomal protein L34	IPB000271 15.87 9.78e-09 161-198 IPB000492B 5.26 9.90e-09 161-195 IPB000221 5.48 1.00e-08 126-152
1998	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.91e-11 52-89 IPB003006B 20.23 3.52e-10 155-192 IPB003006B 20.23 1.69e-09 250-287 IPB003006B 20.23 4.81e-09 437-474
1998	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 5.85e-09 59-82
1999	IPB000897	GTP-binding signal recognition particle (SRP54) domain	IPB000897A 9.15 8.60e-11 313-332
2000	IPB001140	ABC transporter transmembrane	IPB001140A 21.73 2.00e-19 107-153

TABLE 3B

		region	IPB001140B 15.62 4.44e-10 222-260
2000	IPB000795	GTP-binding elongation factor	IPB000795A 10.67 7.88e-10 120-135
2000	PR00326	GTP1/OBG GTP-binding protein family signature I	PR00326A 8.70 4.49e-09 121-141
2000	IPB000897	GTP-binding signal recognition particle (SRP54) domain	IPB000897A 9.15 5.57e-09 120-139
2000	IPB001324	Phosphoribulokinase family	IPB001324A 18.12 8.00e-09 117-138
2001	IPB001422	Neuromodulin (GAP-43)	IPB001422C 16.82 5.26e-10 778-813
2001	PR01217	Proline rich extensin signature VII	PR01217G 4.02 7.16e-09 309-334
2001	IPB003134	Repeat in HS1/Cortactin	IPB003134F 15.66 7.29e-09 776-824 PR01217D 4.57 7.49e-09 562-583
2001	IPB000996	Clathrin light chain	IPB000996B 20.25 7.82e-09 752-804
2001	IPB002079	"Gag polyprotein, inner coat protein p12"	IPB002079J 10.53 9.19e-09 779-807
2001	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135A 11.69 9.62e-09 763-817
2001	IPB001084	Microtubule associated Tau protein	IPB001084C 7.66 9.64e-09 375-392
2001	IPB001101	Plectin repeat	IPB001101K 8.53 9.92e-09 96-139
2002	IPB001552	Acyl-CoA dehydrogenase	IPB001552E 22.77 2.46e-19 523-563 IPB001552D 24.88 5.35e-19 432-474 IPB001552C 25.04 7.75e-15 378-418 IPB001552B 18.05 3.43e-12 124-146 IPB001552A 11.25 6.90e-10 97-108
2003	IPB000998	MAM domain	IPB000998D 18.66 1.96e-15 546-569
2003	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.77e-15 253-272
2003	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 2.89e-14 126-141
2003	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 5.00e-14 208-219 IPB000152 8.86 1.00e-13 253-268 IPB000152 8.86 1.82e-13 208-223 IPB001881B 12.28 4.75e-13 126-137
2003	IPB001774	Delta serrate ligand	IPB001774C 18.25 9.13e-13 88-130 IPB000998B 17.20 1.00e-12 428-440
2003	PR00020	MAM domain signature I	PR00020A 20.48 2.88e-11 426-444 IPB000998C 18.63 5.30e-11 483-498 IPB001881B 12.28 8.58e-11 253-264
2003	PR00907	Thrombomodulin signature II	PR00907B 11.50 2.44e-10 160-176
2003	IPB000561	EGF-like domain	IPB000561 4.89 3.25e-10 97-105
2003	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 5.35e-10 258-268 IPB000033B 7.05 5.97e-09 213-223
2003	IPB000167	Dehydrin	IPB000167A 8.58 7.14e-09 340-367
2003	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 9.79e-09 175-195
2004	IPB001258	NHL repeat	IPB001258B 28.61 4.30e-17 102-136 IPB001258B 28.61 7.00e-17 8-42 IPB001258B 28.61 5.60e-11 55-89
2005	IPB000198	RhoGAP domain	IPB000198C 16.49 8.31e-16 952-969 IPB000198B 12.47 9.10e-15 862-879
2005	IPB002219	Phorbol esters/diacylglycerol binding domain	IPB002219B 12.53 3.89e-11 753-768 IPB000198A 15.95 9.61e-10 810-826
2005	IPB002551	Coronavirus S1 glycoprotein	IPB002551J 18.56 3.60e-09 499-540
2005	IPB001369	Purine and other phosphorylases family 2	IPB001369C 24.81 4.27e-09 65-105
2005	IPB003351	Dishevelled specific domain	IPB003351C 13.82 7.24e-09 1054-1093
2007	PR01303	Plasmodium circumsporozoite protein signature IV	PR01303D 10.57 9.21e-10 5-22
2008	IPB003164	Alpha adaptin carboxyl-terminal domain	IPB003164L 9.84 1.00e-40 48-82 IPB003164N 8.78 1.00e-40 184-222

TABLE 3B

			IPB003164Q 13.71 1.00e-40 285-319 IPB003164S 13.40 1.00e-40 353-394 IPB003164R 10.50 2.35e-38 320-352 IPB003164O 13.89 8.62e-35 223-255 IPB003164P 12.26 7.65e-33 256-284 IPB003164M 10.25 5.18e-31 107-138 IPB003164T 10.57 4.86e-25 395-414
2013	IPB001359	Synapsin	IPB001359H 22.58 2.75e-09 14-64 IPB001359H 22.58 3.62e-09 40-90
2015	PR00456	Ribosomal protein P2 signature V	PR00456E 3.08 5.71e-09 22-36
2016	IPB003134	Repeat in HS1/Cortactin	IPB003134F 15.66 1.48e-09 145-193
2017	PR01297	Colicin lysis protein signature I	PR01297A 6.60 6.02e-09 16-29
2018	PR00205	Cadherin signature IV	PR00205D 12.22 3.25e-16 37-56 PR00205G 13.05 1.37e-13 40-57 PR00205F 19.57 3.10e-13 6-32 PR00205C 13.59 6.62e-09 23-35
2020	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862C 26.48 8.94e-09 113-161
2021	IPB001909	KRAB box	IPB001909 17.37 8.65e-30 56-90
2022	IPB001909	KRAB box	IPB001909 17.37 8.65e-30 56-90
2024	IPB000560	Histidine acid phosphatase	IPB000560 17.02 1.00e-16 35-57
2026	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.00e-24 545-570
2026	IPB001909	KRAB box	IPB001909 17.37 2.86e-21 134-168 IPB000822 14.67 2.29e-17 573-598 IPB000822 14.67 3.57e-17 487-512 IPB000822 14.67 2.50e-13 515-540
2026	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 4.32e-11 570-583 PR00048A 9.94 5.26e-11 484-497 PR00048A 9.94 9.53e-11 542-555 PR00048B 5.52 1.00e-10 558-567 PR00048A 9.94 3.86e-10 512-525
2026	IPB001012	UBX domain	IPB001012A 12.95 7.00e-10 297-312
2026	IPB001580	Calreticulin family	IPB001580F 2.93 1.00e-09 305-314 PR00048B 5.52 6.50e-09 500-509
2026	PR01073	Presenilin 1 signature III	PR01073C 1.45 6.62e-09 300-311
2026	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 9.73e-09 298-322
2029	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 59-86
2029	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 72-86 IPB001599K 8.15 1.46e-10 29-40
2031	PR00014	Fibronectin type III repeat signature IV	PR00014D 15.12 5.26e-10 17-31
2032	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 6.85e-13 118-132
2032	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 7.14e-11 27-40 PR00019B 11.42 8.09e-09 24-37
2033	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 6.85e-13 118-132
2033	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 7.14e-11 27-40 PR00019B 11.42 8.09e-09 24-37
2034	IPB000203	GPS domain	IPB000203A 18.40 9.25e-20 991-1021 IPB000203B 13.98 8.88e-15 1111-1132
2034	IPB000832	G-protein coupled receptors family 2 (secretin-like)	IPB000832C 19.53 9.46e-13 1111-1140
2034	PR00249	Secretin-like GPCR superfamily signature III	PR00249C 15.44 1.73e-10 1113-1136 IPB000832G 15.17 7.81e-09 1281-1306
2035	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 3.45e-21 51-76

435

TABLE 3B

			IPB000822 14.67 4.00e-19 79-104 IPB000822 14.67 3.40e-16 23-48
2035	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 6.54e-14 20-33
2035	IPB001275	DM DNA binding domain	IPB001275 19.17 8.05e-14 11-50 IPB001275 19.17 2.14e-13 39-78 PR00048B 5.52 4.00e-11 92-101 PR00048A 9.94 6.21e-11 76-89 PR00048B 5.52 6.25e-11 64-73 PR00048A 9.94 5.09e-10 104-117 PR00048B 5.52 2.00e-09 8-17 IPB001275 19.17 4.53e-09 67-106 PR00048A 9.94 8.12e-09 48-61
2035	PR00995	36kDa capillovirus serine protease (S35) signature VI	PR00995F 16.50 9.73e-09 1-19
2038	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 8.71e-10 8-22 PR00049D 0.00 9.43e-10 9-23
2038	IPB003861	E4 protein	IPB003861B 9.06 1.98e-09 17-31 PR00049D 0.00 2.37e-09 12-26 PR00049D 0.00 2.53e-09 11-25 PR00049D 0.00 4.36e-09 10-24
2038	IPB002999	Tudor domain	IPB002999B 7.50 7.55e-09 13-21 IPB002999B 7.50 7.55e-09 14-22 IPB002999B 7.50 8.36e-09 11-19
2038	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 8.88e-09 199-224
2039	IPB001310	HIT (Histidine triad) family	IPB001310A 18.76 3.25e-18 197-227 IPB001310B 21.00 2.93e-12 261-287
2039	PR00332	Histidine triad family signature II	PR00332B 14.02 6.26e-10 209-227
2039	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 2.13e-09 339-364
2040	IPB001310	HIT (Histidine triad) family	IPB001310A 18.76 3.25e-18 197-227 IPB001310B 21.00 2.93e-12 261-287
2040	PR00332	Histidine triad family signature II	PR00332B 14.02 6.26e-10 209-227
2040	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 2.13e-09 339-364
2041	IPB001310	HIT (Histidine triad) family	IPB001310A 18.76 3.25e-18 197-227 IPB001310B 21.00 2.93e-12 261-287
2041	PR00332	Histidine triad family signature II	PR00332B 14.02 6.26e-10 209-227
2041	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 2.13e-09 339-364
2042	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 4.52e-10 102-126 IPB000135D 2.13 9.71e-10 104-128 IPB000135D 2.13 9.90e-10 101-125 IPB000135D 2.13 3.18e-09 105-129 IPB000135D 2.13 9.55e-09 103-127
2043	PR00074	Protein-lysine 6-oxidase precursor signature VIII	PR00074H 17.29 8.11e-19 264-283 PR00074E 11.34 3.88e-16 193-213 PR00074F 11.47 6.65e-16 217-238 PR00074B 7.56 4.98e-12 126-150
2043	IPB001695	Lysyl oxidase	IPB001695E 9.12 5.70e-12 110-151 PR00074D 21.66 2.94e-10 171-192
2043	PR00258	Speract receptor signature I	PR00258A 13.56 3.70e-10 5-21 PR00258C 9.05 4.95e-10 43-53 PR00258D 14.29 6.29e-10 76-90 IPB001695F 11.10 6.24e-09 151-179
2046	PR01254	Prostaglandin D synthase signature II	PR01254B 12.05 1.17e-09 339-349
2048	IPB000374	Phosphatidate cytidyltransferase	IPB000374B 15.86 2.06e-27 375-402 IPB000374A 12.59 3.65e-16 271-283
2049	PR00320	G protein beta WD-40 repeat signature I	PR00320A 13.15 7.95e-11 118-132 PR00320B 12.82 2.08e-10 118-132 PR00320C 12.32 4.33e-09 118-132

436

TABLE 3B

2052	PR01446	Claudin-8 signature III	PR01446C 9.62 2.27e-09 119-131
2053	IPB002884	Proprotein convertase P-domain	IPB002884B 15.69 6.33e-09 114-131
2054	IPB000361	Hypothetical hesB/yadR/yfhF family	IPB000361B 19.14 3.08e-19 122-153 IPB000361A 17.83 2.71e-16 73-93
2055	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 9.28e-10 133-170
2055	IPB000920	Myelin P0 protein	IPB000920C 15.78 3.92e-09 161-213
2055	PR00213	Myelin P0 protein signature V	PR00213E 5.51 8.97e-09 179-203
2058	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 3.17e-17 27-79 IPB001442A 26.12 3.60e-17 33-85 IPB001442A 26.12 1.21e-16 39-91
2058	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 2.19e-16 35-88 IPB000885A 11.46 5.06e-16 40-77 IPB001442A 26.12 6.02e-16 30-82 IPB000885B 19.15 3.65e-15 44-97 IPB000885B 19.15 4.39e-15 26-79 IPB000885B 19.15 4.49e-15 32-85 IPB001442A 26.12 9.29e-15 24-76
2058	PR00453	Von Willebrand factor type A domain signature I	PR00453A 11.78 1.75e-14 107-124 IPB000885A 11.46 2.29e-14 43-80 IPB000885A 11.46 3.92e-14 52-89 IPB000885B 19.15 6.97e-14 29-82 IPB001442A 26.12 7.65e-14 42-94 IPB001442A 26.12 8.63e-14 45-97 IPB001442A 26.12 1.00e-13 36-88 IPB000885A 11.46 2.89e-13 37-74 IPB000885A 11.46 6.33e-13 49-86 IPB000885B 19.15 7.07e-13 38-91 IPB000885B 19.15 7.46e-13 41-94
2058	IPB001073	Complement C1q protein	IPB001073A 22.14 1.72e-12 45-79 IPB000885A 11.46 5.93e-12 55-92 IPB000885A 11.46 6.04e-12 46-83 IPB001073A 22.14 7.48e-12 48-82 IPB000885B 19.15 7.84e-12 23-76 IPB000885B 19.15 8.88e-12 47-100 IPB001442B 12.38 9.85e-12 61-81
2059	IPB001541	SUR2-type hydroxylase/desaturase catalytic domain	IPB001541A 12.30 5.50e-11 40-52 IPB001541B 11.65 4.86e-09 127-136
2060	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.19e-09 134-171
2061	PR00918	Calicivirus non-structural polyprotein family signature I	PR00918A 13.81 3.59e-12 37-57
2061	IPB002078	Sigma-54 factor interaction protein family	IPB002078A 20.43 6.31e-10 43-77
2061	PR00364	Disease resistance protein signature I	PR00364A 8.29 7.11e-10 42-57
2061	IPB000765	GTP1/OBG family	IPB000765 26.91 7.67e-10 41-84
2061	PR00094	Adenylate kinase signature I	PR00094A 9.62 2.43e-09 44-57
2061	PR00830	Endopeptidase La (Lon) serine protease (S16) signature I	PR00830A 8.52 4.50e-09 47-66
2067	PR00874	Fungi-IV metallothionein signature III	PR00874C 4.37 6.50e-09 7-21
2071	PR01539	Interleukin-1 receptor type II precursor signature IX	PR01539I 14.65 9.06e-09 223-246
2074	IPB001284	Ribosomal protein L34e	IPB001284A 18.97 3.48e-31 15-50 IPB001284B 26.99 1.41e-28 53-85
2074	PR01250	Ribosomal protein L34 signature IV	PR01250D 13.87 2.69e-23 73-95 PR01250B 13.36 7.92e-17 33-50

437

TABLE 3B

			PR01250A 11.25 2.25e-13 20-33 PR01250C 9.53 4.52e-12 53-63 IPB001284B 26.99 3.75e-09 82-114
2076	IPB000171	Bacterial-type phytoene dehydrogenase	IPB000171E 7.19 8.20e-09 294-304
2077	IPB001774	Delta serrate ligand	IPB001774D 19.23 5.91e-09 50-96
2077	IPB000034	Laminin B	IPB000034C 12.97 7.31e-09 84-102
2077	IPB000561	EGF-like domain	IPB000561 4.89 8.07e-09 84-92
2078	IPB001774	Delta serrate ligand	IPB001774D 19.23 5.91e-09 50-96
2078	IPB000034	Laminin B	IPB000034C 12.97 7.31e-09 84-102
2078	IPB000561	EGF-like domain	IPB000561 4.89 8.07e-09 84-92
2079	IPB001774	Delta serrate ligand	IPB001774D 19.23 5.91e-09 50-96
2079	IPB000034	Laminin B	IPB000034C 12.97 7.31e-09 84-102
2079	IPB000561	EGF-like domain	IPB000561 4.89 8.07e-09 84-92
2080	PR00436	Interleukin-8 signature I	PR00436A 15.20 9.36e-10 14-37
2081	IPB001187	Tissue Factor (TF)	IPB001187G 15.20 7.00e-10 33-69
2081	IPB001073	Complement C1q protein	IPB001073A 22.14 2.69e-09 146-180
2081	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 6.03e-09 205-219 PR00049D 0.00 6.34e-09 207-221 PR00049D 0.00 7.41e-09 203-217
2081	PR00499	Neutrophil cytosol factor 2 signature I	PR00499A 7.48 7.60e-09 791-808
2081	IPB001359	Synapsin	IPB001359H 22.58 8.08e-09 772-822
2081	IPB003036	Gag P30 core shell protein	IPB003036C 11.53 9.63e-09 155-171
2082	IPB001039	"Major histocompatibility complex protein, Class I"	IPB001039B 27.55 3.01e-09 103-154
2083	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 148-185 IPB003006B 20.23 9.14e-12 441-478 IPB003006B 20.23 1.00e-11 248-285
2083	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.23e-11 547-570 IPB003006B 20.23 6.40e-10 54-91 IPB003006B 20.23 9.64e-10 540-577 IPB003006B 20.23 8.62e-09 346-383 PR01536C 19.92 9.19e-09 155-178
2084	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 148-185 IPB003006B 20.23 9.14e-12 441-478 IPB003006B 20.23 1.00e-11 248-285
2084	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.23e-11 547-570 IPB003006B 20.23 6.40e-10 54-91 IPB003006B 20.23 9.64e-10 540-577 IPB003006B 20.23 8.62e-09 346-383 PR01536C 19.92 9.19e-09 155-178
2085	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 148-185 IPB003006B 20.23 9.14e-12 441-478 IPB003006B 20.23 1.00e-11 248-285
2085	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.23e-11 547-570 IPB003006B 20.23 6.40e-10 54-91 IPB003006B 20.23 9.64e-10 540-577 IPB003006B 20.23 8.62e-09 346-383 PR01536C 19.92 9.19e-09 155-178
2086	IPB002117	p53 tumor antigen	IPB002117A 9.71 5.50e-15 13-23
2087	IPB000074	Apolipoprotein A1/A4/E	IPB000074B 29.17 7.49e-10 117-170 IPB000074B 29.17 8.75e-10 95-148 IPB000074B 29.17 9.20e-10 62-115 IPB000074C 22.23 2.62e-09 90-127 IPB000074C 22.23 4.35e-09 112-149 IPB000074B 29.17 8.48e-09 201-254

TABLE 3B

2088	IPB000074	Apolipoprotein A1/A4/E	IPB000074B 29.17 7.49e-10 117-170 IPB000074B 29.17 8.75e-10 95-148 IPB000074B 29.17 9.20e-10 62-115 IPB000074C 22.23 2.62e-09 90-127 IPB000074C 22.23 4.35e-09 112-149 IPB000074B 29.17 8.48e-09 201-254
2090	IPB001211	Phospholipase A2	IPB001211B 17.16 3.12e-31 49-76
2090	PR00389	Phospholipase A2 signature III	PR00389C 17.85 2.50e-20 61-79 PR00389B 10.67 6.91e-16 42-60 IPB001211D 11.66 5.50e-14 109-124 PR00389E 13.06 8.20e-14 109-125 IPB001211C 14.62 1.56e-11 84-102
2091	PR01217	Proline rich extensin signature VI	PR01217F 4.24 8.40e-09 65-82
2092	IPB001354	Mandelate racemase/muconate lactonizing enzyme family	IPB001354C 32.55 1.00e-24 255-296 IPB001354D 32.92 2.07e-18 343-388 IPB001354B 18.16 3.91e-18 132-158
2094	IPB000222	Protein phosphatase 2C subfamily	IPB000222F 19.87 4.94e-15 285-305 IPB000222E 14.28 6.33e-15 257-275 IPB000222G 9.17 1.95e-12 311-324 IPB000222C 6.84 2.08e-12 176-185 IPB000222H 9.33 7.97e-12 347-359 IPB000222B 15.80 2.86e-10 144-154 IPB000222D 11.74 2.74e-09 215-232 IPB000222I 8.91 4.72e-09 408-417
2095	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 4.71e-15 107-122 IPB000152 8.86 1.47e-14 44-59
2095	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 1.47e-11 107-118
2095	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 4.96e-11 49-59 IPB001881B 12.28 6.68e-11 44-55
2095	PR00010	Type II EGF-like signature III	PR00010C 6.98 7.10e-10 49-59 PR00010C 6.98 7.68e-10 112-122 IPB001881B 12.28 2.57e-09 5-16 IPB000033B 7.05 3.13e-09 112-122
2095	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 5.71e-09 107-126
2096	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 218-234
2096	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 160-171 PR00245D 9.34 1.53e-13 278-287 PR00245E 8.96 6.81e-12 325-336 PR00245B 13.73 1.00e-10 171-183 IPB000276D 9.40 3.08e-09 324-340
2096	PR00237	Rhodopsin-like GPCR superfamily signature V	PR00237E 13.03 3.83e-09 241-264
2096	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.17e-09 93-105 PR00237C 14.77 5.91e-09 146-168
2096	PR00896	Vasopressin receptor signature II	PR00896B 9.36 7.23e-09 97-108 PR00237G 19.23 1.00e-08 314-340
2097	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 218-234
2097	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 160-171 PR00245D 9.34 1.53e-13 278-287 PR00245E 8.96 6.81e-12 325-336 PR00245B 13.73 1.00e-10 171-183 IPB000276D 9.40 3.08e-09 324-340
2097	PR00237	Rhodopsin-like GPCR superfamily signature V	PR00237E 13.03 3.83e-09 241-264
2097	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.17e-09 93-105 PR00237C 14.77 5.91e-09 146-168
2097	PR00896	Vasopressin receptor signature II	PR00896B 9.36 7.23e-09 97-108

TABLE 3B

			PR00237G 19.23 1.00e-08 314-340
2098	IPB001169	"Integrin beta, C-terminus"	IPB001169J 7.42 4.63e-10 49-62
2098	PR01186	Integrin beta subunit signature XI	PR01186K 7.39 7.27e-10 49-62 PR01186K 7.39 9.75e-09 15-28
2102	PR00193	Myosin heavy chain signature III	PR00193C 11.66 9.77e-24 126-153
2102	IPB000857	Core domain in kinesin and myosin motors	IPB000857C 10.82 4.84e-19 124-146 PR00193B 12.36 6.81e-18 74-99 IPB000857D 12.93 7.64e-12 153-191 PR00193A 14.87 8.50e-12 14-33 IPB000857B 11.35 1.00e-10 55-101
2102	PR00364	Disease resistance protein signature I	PR00364A 8.29 4.86e-09 76-91
2103	PR00193	Myosin heavy chain signature III	PR00193C 11.66 9.77e-24 126-153
2103	IPB000857	Core domain in kinesin and myosin motors	IPB000857C 10.82 4.84e-19 124-146 PR00193B 12.36 6.81e-18 74-99 IPB000857D 12.93 7.64e-12 153-191 PR00193A 14.87 8.50e-12 14-33 IPB000857B 11.35 1.00e-10 55-101
2103	PR00364	Disease resistance protein signature I	PR00364A 8.29 4.86e-09 76-91
2105	IPB002350	Kazal-type serine protease inhibitor family	IPB002350 31.78 2.86e-18 77-117
2105	IPB000716	Thyroglobulin type-1 repeat	IPB000716C 17.62 2.88e-18 274-292 IPB000716D 15.49 7.16e-15 296-310
2109	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 5.50e-13 45-59
2111	IPB000221	Protamine P1	IPB000221 5.48 3.08e-09 3-29
2112	PR01415	Ankyrin repeat signature II	PR01415B 10.23 5.88e-09 26-38
2113	IPB000416	Outer Capsid protein VP4 (Hemagglutinin)	IPB000416P 15.37 7.00e-09 188-226
2114	IPB000416	Outer Capsid protein VP4 (Hemagglutinin)	IPB000416P 15.37 7.00e-09 188-226
2115	IPB000998	MAM domain	IPB000998C 18.63 1.95e-12 17-32
2115	PR00020	MAM domain signature III	PR00020C 12.01 8.12e-10 16-27 IPB000998D 18.66 9.61e-10 82-105
2116	IPB000998	MAM domain	IPB000998C 18.63 1.95e-12 17-32
2116	PR00020	MAM domain signature III	PR00020C 12.01 8.12e-10 16-27 IPB000998D 18.66 9.61e-10 82-105
2118	IPB002642	Lysophospholipase catalytic domain	IPB002642E 18.19 6.91e-10 86-111
2119	IPB002642	Lysophospholipase catalytic domain	IPB002642E 18.19 6.91e-10 86-111
2120	IPB000817	Prion protein	IPB000817A 8.34 7.73e-10 255-297
2120	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 7.26e-09 262-314
2122	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.43e-13 72-109
2122	IPB003531	Short hematopoietin receptor family 1	IPB003531C 15.87 9.38e-11 318-335
2123	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.43e-13 72-109
2123	IPB003531	Short hematopoietin receptor family 1	IPB003531C 15.87 9.38e-11 318-335
2124	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.43e-13 72-109
2124	IPB003531	Short hematopoietin receptor family 1	IPB003531C 15.87 9.38e-11 318-335
2125	IPB000008	C2 domain	IPB000008C 23.37 7.94e-25 109-148
2125	PR00360	C2 domain signature I	PR00360A 15.18 1.60e-13 107-119
2125	PR00399	Synaptotagmin signature II	PR00399B 14.30 1.69e-12 94-107 IPB000008D 14.83 3.86e-11 164-182

440

TABLE 3B

			PR00360B 11.64 5.94e-11 136-149 PR00399C 15.89 4.98e-10 151-166 PR00399D 12.72 6.33e-10 171-181 PR00399A 15.05 8.65e-09 79-94
2126	IPB002870	Reprolysin family propeptide	IPB002870B 24.73 3.78e-14 142-180
2126	IPB001670	Iron-containing alcohol dehydrogenase	IPB001670D 13.90 5.50e-09 158-173
2130	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 7.53e-26 8-60
2130	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 4.52e-24 1-54 IPB000885B 19.15 2.38e-23 19-72 IPB001442A 26.12 8.04e-23 11-63 IPB001442A 26.12 8.83e-23 20-72 IPB000885B 19.15 2.32e-22 4-57 IPB001442A 26.12 2.93e-22 5-57 IPB001442A 26.12 5.37e-22 17-69
2130	PR00258	Speract receptor signature I	PR00258A 13.56 6.32e-16 87-103 IPB001442A 26.12 7.91e-16 26-78 IPB000885A 11.46 1.49e-15 33-70 IPB000885A 11.46 5.74e-15 24-61 IPB000885B 19.15 5.98e-15 28-81 IPB000885A 11.46 8.30e-15 9-46 IPB000885A 11.46 2.99e-14 30-67 IPB000885B 19.15 4.13e-14 31-84
2130	IPB001073	Complement C1q protein	IPB001073A 22.14 8.40e-14 17-51 IPB000885A 11.46 8.60e-14 39-76 IPB000885B 19.15 2.17e-13 34-87 IPB001073A 22.14 7.89e-13 23-57 PR00258B 7.94 8.42e-13 106-117 IPB001442A 26.12 2.17e-12 35-87 IPB001442B 12.38 2.98e-12 24-44 IPB001442B 12.38 5.58e-12 21-41 IPB001073A 22.14 6.94e-12 20-54 IPB001073A 22.14 8.38e-12 11-45 IPB001442A 26.12 8.47e-12 32-84 IPB001442B 12.38 8.47e-12 12-32 IPB001073A 22.14 8.74e-12 29-63 IPB001442B 12.38 9.69e-12 15-35 IPB001442B 12.38 1.71e-11 51-71 IPB001442B 12.38 2.86e-11 9-29 IPB001073A 22.14 3.83e-11 14-48 IPB000885B 19.15 5.90e-11 40-93 IPB001442B 12.38 8.86e-11 6-26 IPB001073A 22.14 9.17e-11 44-78 IPB001073A 22.14 9.50e-11 2-36 IPB001073A 22.14 1.15e-10 8-42 IPB001073A 22.14 2.83e-10 26-60
2130	IPB000817	Prion protein	IPB000817A 8.34 2.88e-10 1-43 IPB000885B 19.15 4.09e-10 37-90 IPB000885A 11.46 4.23e-10 42-79 IPB001073A 22.14 4.81e-10 47-81 IPB001073A 22.14 5.12e-10 50-84 IPB001073A 22.14 6.03e-10 5-39 IPB001442A 26.12 9.26e-10 38-90 IPB001442B 12.38 1.24e-09 18-38 IPB001073A 22.14 2.13e-09 41-75 IPB001442B 12.38 2.70e-09 3-23 IPB001442B 12.38 4.65e-09 45-65

441

TABLE 3B

			IPB001442B 12.38 5.62e-09 27-47 IPB000885A 11.46 5.87e-09 45-82 IPB001442B 12.38 6.84e-09 48-68 IPB001073A 22.14 9.30e-09 38-72
2131	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 7.53e-26 8-60
2131	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 4.52e-24 1-54 IPB000885B 19.15 2.38e-23 19-72 IPB001442A 26.12 8.04e-23 11-63 IPB001442A 26.12 8.83e-23 20-72 IPB000885B 19.15 2.32e-22 4-57 IPB001442A 26.12 2.93e-22 5-57 IPB001442A 26.12 5.37e-22 17-69
2131	PR00258	Speract receptor signature I	PR00258A 13.56 6.32e-16 87-103 IPB001442A 26.12 7.91e-16 26-78 IPB000885A 11.46 1.49e-15 33-70 IPB000885A 11.46 5.74e-15 24-61 IPB000885B 19.15 5.98e-15 28-81 IPB000885A 11.46 8.30e-15 9-46 IPB000885A 11.46 2.99e-14 30-67 IPB000885B 19.15 4.13e-14 31-84
2131	IPB001073	Complement C1q protein	IPB001073A 22.14 8.40e-14 17-51 IPB000885A 11.46 8.60e-14 39-76 IPB000885B 19.15 2.17e-13 34-87 IPB001073A 22.14 7.89e-13 23-57 PR00258B 7.94 8.42e-13 106-117 IPB001442A 26.12 2.17e-12 35-87 IPB001442B 12.38 2.98e-12 24-44 IPB001442B 12.38 5.58e-12 21-41 IPB001073A 22.14 6.94e-12 20-54 IPB001073A 22.14 8.38e-12 11-45 IPB001442A 26.12 8.47e-12 32-84 IPB001442B 12.38 8.47e-12 12-32 IPB001073A 22.14 8.74e-12 29-63 IPB001442B 12.38 9.69e-12 15-35 IPB001442B 12.38 1.71e-11 51-71 IPB001442B 12.38 2.86e-11 9-29 IPB001073A 22.14 3.83e-11 14-48 IPB000885B 19.15 5.90e-11 40-93 IPB001442B 12.38 8.86e-11 6-26 IPB001073A 22.14 9.17e-11 44-78 IPB001073A 22.14 9.50e-11 2-36 IPB001073A 22.14 1.15e-10 8-42 IPB001073A 22.14 2.83e-10 26-60
2131	IPB000817	Prion protein	IPB000817A 8.34 2.88e-10 1-43 IPB000885B 19.15 4.09e-10 37-90 IPB000885A 11.46 4.23e-10 42-79 IPB001073A 22.14 4.81e-10 47-81 IPB001073A 22.14 5.12e-10 50-84 IPB001073A 22.14 6.03e-10 5-39 IPB001442A 26.12 9.26e-10 38-90 IPB001442B 12.38 1.24e-09 18-38 IPB001073A 22.14 2.13e-09 41-75 IPB001442B 12.38 2.70e-09 3-23 IPB001442B 12.38 4.65e-09 45-65 IPB001442B 12.38 5.62e-09 27-47 IPB000885A 11.46 5.87e-09 45-82 IPB001442B 12.38 6.84e-09 48-68

TABLE 3B

			IPB001073A 22.14 9.30e-09 38-72
2132	IPB000237	GRIP domain	IPB000237B 30.66 3.22e-10 427-477
2133	IPB001909	KRAB box	IPB001909 17.37 6.50e-34 63-97
2133	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 8.20e-22 354-379 IPB000822 14.67 5.09e-21 438-463 IPB000822 14.67 5.50e-20 606-631 IPB000822 14.67 7.00e-20 578-603 IPB000822 14.67 3.25e-19 522-547 IPB000822 14.67 4.00e-19 326-351 IPB000822 14.67 7.00e-19 410-435 IPB000822 14.67 4.46e-18 494-519 IPB000822 14.67 6.14e-17 382-407 IPB000822 14.67 3.40e-16 550-575 IPB000822 14.67 4.00e-16 466-491
2133	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 5.85e-14 547-560 PR00048A 9.94 8.07e-13 351-364 PR00048A 9.94 3.12e-12 519-532 PR00048A 9.94 4.71e-12 379-392 PR00048A 9.94 4.71e-12 463-476 PR00048B 5.52 7.00e-12 619-628
2133	IPB001275	DM DNA binding domain	IPB001275 19.17 7.04e-12 398-437 PR00048A 9.94 7.88e-12 631-644 PR00048A 9.94 1.95e-11 603-616 PR00048A 9.94 4.32e-11 575-588 PR00048B 5.52 5.50e-11 451-460 PR00048A 9.94 1.00e-10 323-336 IPB001275 19.17 1.36e-10 426-465 IPB001275 19.17 1.49e-10 482-521 PR00048A 9.94 5.09e-10 435-448 IPB001275 19.17 5.14e-10 510-549
2133	IPB002817	ThiC family	IPB002817H 11.39 5.42e-10 349-364 PR00048A 9.94 5.91e-10 491-504 IPB001275 19.17 8.18e-10 314-353 IPB001275 19.17 9.15e-10 454-493 PR00048B 5.52 9.36e-10 507-516 IPB001275 19.17 9.39e-10 342-381 IPB001275 19.17 9.39e-10 370-409 PR00048B 5.52 2.00e-09 339-348 IPB000822 14.67 2.13e-09 634-659 PR00048B 5.52 2.50e-09 591-600 IPB001275 19.17 2.71e-09 594-633 PR00048B 5.52 3.00e-09 535-544 IPB001275 19.17 3.62e-09 538-577 PR00048A 9.94 4.38e-09 407-420
2133	IPB000306	"FYVE Zn-finger, rabphilin/VPS27/FAB1 type"	IPB000306 8.96 4.71e-09 350-362 PR00048B 5.52 5.50e-09 423-432 IPB000306 8.96 5.76e-09 630-642 IPB000306 8.96 6.03e-09 434-446 PR00048B 5.52 7.00e-09 367-376 IPB002817H 11.39 7.34e-09 433-448 IPB001275 19.17 8.18e-09 566-605
2133	IPB002634	BolA-like protein	IPB002634A 23.30 8.62e-09 375-409
2137	IPB000954	Aminotransferase class-III pyridoxal-phosphate	IPB000954B 21.02 9.38e-21 191-230 IPB000954D 13.61 5.74e-17 277-295 IPB000954C 12.88 9.44e-14 240-255
2138	IPB000954	Aminotransferase class-III pyridoxal-phosphate	IPB000954B 21.02 9.38e-21 191-230 IPB000954D 13.61 5.74e-17 277-295 IPB000954C 12.88 9.44e-14 240-255

TABLE 3B

2139	IPB001254	"Serine proteases, trypsin family"	IPB001254A 9.98 6.14e-15 33-49
2139	PR00722	Chymotrypsin serine protease family (S1) signature I	PR00722A 12.06 4.54e-14 34-49
2139	IPB000001	Kringle	IPB000001D 11.31 7.56e-12 33-49
2139	IPB000177	Apple domain	IPB000177K 13.19 2.57e-10 35-67 PR00722B 12.69 6.85e-10 90-104
2142	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 3.89e-11 10-25 IPB000152 8.86 4.86e-11 128-143
2142	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 7.63e-11 10-21
2142	PR00010	Type II EGF-like signature III	PR00010C 6.98 2.74e-10 133-143
2142	IPB002899	EB module	IPB002899B 11.81 5.59e-10 116-128 IPB002899B 11.81 5.59e-10 157-169 IPB001881B 12.28 6.57e-10 128-139 IPB001881B 12.28 8.29e-10 169-180 IPB001881A 8.72 9.36e-10 41-50 IPB000152 8.86 9.72e-10 169-184
2142	IPB001862	Membrane attack complex components/perforin/complement C9	IPB001862F 29.39 9.81e-10 26-73 IPB001862F 29.39 1.28e-09 102-149
2142	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 5.03e-09 133-143 PR00010A 12.91 7.27e-09 37-48
2142	IPB000561	EGF-like domain	IPB000561 4.89 7.43e-09 96-104 IPB000561 4.89 7.43e-09 137-145
2144	IPB000608	Ubiquitin-conjugating enzymes	IPB000608 27.71 7.95e-12 72-116
2146	IPB002181	Fibrinogen beta and gamma chains C-terminal globular domain	IPB002181B 20.16 7.49e-24 30-66 IPB002181D 29.18 7.32e-15 92-132 IPB002181C 15.87 2.64e-10 71-83
2147	IPB002181	Fibrinogen beta and gamma chains C-terminal globular domain	IPB002181B 20.16 7.49e-24 30-66 IPB002181D 29.18 7.32e-15 92-132 IPB002181C 15.87 2.64e-10 71-83
2148	IPB002181	Fibrinogen beta and gamma chains C-terminal globular domain	IPB002181B 20.16 7.49e-24 30-66 IPB002181D 29.18 7.32e-15 92-132 IPB002181C 15.87 2.64e-10 71-83
2151	IPB002027	Amino acid permease	IPB002027D 22.00 4.13e-25 248-287 IPB002027C 19.67 2.74e-22 167-205 IPB002027B 12.67 7.97e-12 103-122
2159	PR00503	Bromodomain signature IV	PR00503D 19.24 3.57e-21 432-451
2159	IPB001487	Bromodomain	IPB001487B 17.44 2.13e-19 423-444 PR00503B 10.44 4.37e-19 105-121 IPB001487A 11.44 5.20e-19 106-124 PR00503C 19.09 4.00e-17 121-139 IPB001487A 11.44 9.53e-16 399-417 PR00503A 14.57 4.00e-14 89-102 PR00503B 10.44 8.64e-14 398-414 PR00503D 19.24 9.25e-13 139-158 IPB001487B 17.44 1.58e-12 130-151 PR00503C 19.09 6.70e-11 414-432
2159	IPB001505	"Cu(A) centre of cytochrome c oxidase, subunit II and nitrous oxide reductase"	IPB001505B 15.93 5.94e-10 417-466 IPB001505A 18.04 1.17e-09 104-151
2159	IPB003351	Dishevelled specific domain	IPB003351C 13.82 5.13e-09 496-535 PR00503A 14.57 6.81e-09 382-395
2159	PR01217	Proline rich extensin signature IV	PR01217D 4.57 7.49e-09 250-271
2159	PR01503	Treacher Collins syndrome protein Treacle signature II	PR01503B 3.77 7.64e-09 714-727
2159	IPB000574	Tymovirus coat protein	IPB000574A 32.18 7.78e-09 265-312
2159	PR00910	Luteovirus ORF6 protein signature I	PR00910A 2.74 8.07e-09 266-278
2159	IPB001359	Synapsin	IPB001359H 22.58 8.51e-09 204-254

TABLE 3B

2159	IPB001978	Troponin	IPB001978B 22.99 9.15e-09 541-572
2160	IPB002862	Protein of unknown function DUF16	IPB002862C 11.30 9.59e-09 60-82
2164	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 5.29e-29 7-48
2164	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 2.80e-19 11-49
2164	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861G 13.73 9.60e-16 13-62
2164	IPB001772	Kinase associated domain 1	IPB001772E 24.88 2.25e-14 69-108
2164	IPB003527	MAP kinase	IPB003527G 17.26 8.86e-14 81-118 IPB001772D 21.67 4.73e-13 18-57 IPB003527D 21.53 4.66e-11 4-45
2164	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 9.65e-10 15-69
2164	IPB000959	POLO box duplicated region	IPB000959D 27.01 2.97e-09 62-114
2165	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 5.29e-29 7-48
2165	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 2.80e-19 11-49
2165	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861G 13.73 9.60e-16 13-62
2165	IPB001772	Kinase associated domain 1	IPB001772E 24.88 2.25e-14 69-108
2165	IPB003527	MAP kinase	IPB003527G 17.26 8.86e-14 81-118 IPB001772D 21.67 4.73e-13 18-57 IPB003527D 21.53 4.66e-11 4-45
2165	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 9.65e-10 15-69
2165	IPB000959	POLO box duplicated region	IPB000959D 27.01 2.97e-09 62-114
2167	IPB001073	Complement C1q protein	IPB001073B 20.88 6.00e-26 147-181 IPB001073A 22.14 4.48e-20 101-135
2167	IPB000885	Fibrillar collagen C-terminal domain	IPB000885B 19.15 9.63e-20 70-123
2167	IPB001442	C-terminal tandem repeated domain in type 4 procollagen	IPB001442A 26.12 4.27e-19 71-123 IPB000885B 19.15 7.48e-19 76-129 IPB000885A 11.46 1.97e-18 78-115 IPB000885A 11.46 2.94e-18 84-121
2167	PR00007	Complement C1Q domain signature III	PR00007C 16.13 3.67e-18 215-236 IPB001442A 26.12 1.11e-17 80-132 PR00007A 20.64 1.84e-17 140-166 IPB001442A 26.12 1.87e-17 86-138 IPB000885B 19.15 5.39e-17 73-126 IPB000885A 11.46 6.96e-17 81-118 IPB000885B 19.15 8.87e-17 67-
2167	IPB000817	Prion protein	IPB000817A 8.34 3.27e-09 67-109 IPB000885A 11.46 3.66e-09 35-72 IPB001442A 26.12 4.13e-09 28-80 IPB000885B 19.15 4.19e-09 42-95 IPB000885A 11.46 4.77e-09 102-139 IPB001442A 26.12 4.83e-09 40-92 IPB001442B 12.38 5.99e-09 53-73 IPB001442A 26.12 6.17e-09 37-89 IPB000885B 19.15 7.55e-09 52-105 IPB001442B 12.38 7.57e-09 87-107 IPB001442B 12.38 8.54e-09 105-125 IPB001073A 22.14 8.59e-09 46-80 IPB000885B 19.15 8.69e-09 94-147 IPB001442B 12.38 9.64e-09 90-110
2169	IPB002360	Involucrin	IPB002360C 15.36 3.06e-14 206-247
2169	PR00209	Alpha/beta gliadin family signature II	PR00209B 4.73 5.94e-12 226-244 IPB002360C 15.36 5.93e-10 215-256 IPB002360C 15.36 2.50e-09 195-236 IPB002360C 15.36 2.50e-09 214-255
2169	IPB001359	Synapsin	IPB001359H 22.58 5.19e-09 220-270 IPB002360C 15.36 5.20e-09 203-244

445

TABLE 3B

			IPB002360C 15.36 5.70e-09 212-253 IPB002360C 15.36 6.10e-09 188-229
2169	IPB003753	"Exonuclease VII, large subunit"	IPB003753F 28.29 7.54e-09 181-231 IPB002360C 15.36 8.80e-09 218-259
2170	IPB002360	Involucrin	IPB002360C 15.36 3.06e-14 206-247
2170	PR00209	Alpha/beta gliadin family signature II	PR00209B 4.73 5.94e-12 226-244 IPB002360C 15.36 5.93e-10 215-256 IPB002360C 15.36 2.50e-09 195-236 IPB002360C 15.36 2.50e-09 214-255
2170	IPB001359	Synapsin	IPB001359H 22.58 5.19e-09 220-270 IPB002360C 15.36 5.20e-09 203-244 IPB002360C 15.36 5.70e-09 212-253 IPB002360C 15.36 6.10e-09 188-229
2170	IPB003753	"Exonuclease VII, large subunit"	IPB003753F 28.29 7.54e-09 181-231 IPB002360C 15.36 8.80e-09 218-259
2172	IPB000483	Leucine rich repeat C-terminal domain	IPB000483 11.18 5.50e-13 45-59
2173	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.83e-11 69-106
2175	PR00457	Animal haem peroxidase signature VII	PR00457G 14.17 4.48e-14 144-164 PR00457H 14.82 5.85e-13 215-229 PR00457F 14.42 6.32e-12 17-27
2176	PR00457	Animal haem peroxidase signature VII	PR00457G 14.17 4.48e-14 144-164 PR00457H 14.82 5.85e-13 215-229 PR00457F 14.42 6.32e-12 17-27
2177	PR00457	Animal haem peroxidase signature VII	PR00457G 14.17 4.48e-14 144-164 PR00457H 14.82 5.85e-13 215-229 PR00457F 14.42 6.32e-12 17-27
2179	IPB002151	Kinesin light chain repeat	IPB002151B 14.23 8.01e-10 259-311
2179	IPB000421	Coagulation factor 5/8 type C domain (FA58C)	IPB000421A 21.21 7.85e-09 62-81
2180	IPB003117	Regulatory subunit of type II PKA R-subunit	IPB003117C 17.01 1.00e-40 189-229 IPB003117D 18.87 1.00e-40 240-280 IPB003117G 17.45 8.50e-33 383-417 IPB003117A 22.23 5.50e-26 66-98 IPB003117E 18.84 5.85e-23 329-357
2180	IPB000595	Cyclic nucleotide-binding domain	IPB000595C 23.31 6.82e-21 363-388
2180	PR00103	cAMP-dependent protein kinase signature II	PR00103B 10.32 7.00e-18 215-229 IPB000595B 15.72 7.50e-18 321-344 IPB003117F 17.26 1.00e-17 365-379 IPB000595B 15.72 4.43e-16 203-226 PR00103A 9.07 7.75e-16 200-214 IPB003117C 17.01 2.96e-15 307-347 IPB003117D 18.87 4.14e-15 364-404 PR00103E 12.91 5.91e-14 397-409 PR00103D 10.18 2.93e-13 376-387 IPB000595C 23.31 4.60e-13 239-264 PR00103C 13.28 1.84e-11 364-373 PR00103D 10.18 2.98e-10 252-263 IPB003117E 18.84 3.57e-10 199-227 IPB003117E 18.84 5.43e-10 317-345 IPB003117F 17.26 1.50e-09 241-255 PR00103A 9.07 8.11e-09 318-332
2181	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478B 6.12 4.94e-09 49-58
2182	IPB000907	Lipoxygenase	IPB000907J 20.31 5.50e-37 499-541 IPB000907G 22.23 1.87e-34 346-388

TABLE 3B

			IPB000907F 21.29 1.00e-28 313-345
2182	PR00467	Mammalian lipoxygenase signature VI	PR00467F 12.25 9.41e-22 393-415
2182	PR00087	Lipoxygenase signature III	PR00087C 13.32 1.39e-21 348-368 IPB000907C 16.09 7.17e-21 195-221 IPB000907I 27.52 7.16e-19 438-491 IPB000907E 15.16 9.21e-18 270-294 PR00467D 17.16 9.57e-17 170-191 IPB000907D 18.70 2.67e-16 236-263 PR00467E 9.17 1.16e-15 267-286 PR00087A 20.06 3.52e-15 310-327 PR00087B 13.69 5.11e-15 328-345 IPB000907B 14.10 2.50e-13 132-147 PR00467A 8.38 3.29e-13 11-28 IPB000907H 18.37 5.86e-13 409-425 PR00467B 14.98 5.88e-12 57-76 PR00467G 16.61 3.37e-11 554-571 IPB000907A 16.20 4.21e-10 94-103
2183	IPB000907	Lipoxygenase	IPB000907J 20.31 5.50e-37 499-541 IPB000907G 22.23 1.87e-34 346-388 IPB000907F 21.29 1.00e-28 313-345
2183	PR00467	Mammalian lipoxygenase signature VI	PR00467F 12.25 9.41e-22 393-415
2183	PR00087	Lipoxygenase signature III	PR00087C 13.32 1.39e-21 348-368 IPB000907C 16.09 7.17e-21 195-221 IPB000907I 27.52 7.16e-19 438-491 IPB000907E 15.16 9.21e-18 270-294 PR00467D 17.16 9.57e-17 170-191 IPB000907D 18.70 2.67e-16 236-263 PR00467E 9.17 1.16e-15 267-286 PR00087A 20.06 3.52e-15 310-327 PR00087B 13.69 5.11e-15 328-345 IPB000907B 14.10 2.50e-13 132-147 PR00467A 8.38 3.29e-13 11-28 IPB000907H 18.37 5.86e-13 409-425 PR00467B 14.98 5.88e-12 57-76 PR00467G 16.61 3.37e-11 554-571 IPB000907A 16.20 4.21e-10 94-103
2184	IPB000907	Lipoxygenase	IPB000907J 20.31 5.50e-37 499-541 IPB000907G 22.23 1.87e-34 346-388 IPB000907F 21.29 1.00e-28 313-345
2184	PR00467	Mammalian lipoxygenase signature VI	PR00467F 12.25 9.41e-22 393-415
2184	PR00087	Lipoxygenase signature III	PR00087C 13.32 1.39e-21 348-368 IPB000907C 16.09 7.17e-21 195-221 IPB000907I 27.52 7.16e-19 438-491 IPB000907E 15.16 9.21e-18 270-294 PR00467D 17.16 9.57e-17 170-191 IPB000907D 18.70 2.67e-16 236-263 PR00467E 9.17 1.16e-15 267-286 PR00087A 20.06 3.52e-15 310-327 PR00087B 13.69 5.11e-15 328-345 IPB000907B 14.10 2.50e-13 132-147 PR00467A 8.38 3.29e-13 11-28 IPB000907H 18.37 5.86e-13 409-425 PR00467B 14.98 5.88e-12 57-76 PR00467G 16.61 3.37e-11 554-571 IPB000907A 16.20 4.21e-10 94-103

TABLE 3B

2193	IPB001774	Delta serrate ligand	IPB001774C 18.25 1.71e-31 37-79 IPB001774D 19.23 3.32e-25 83-129
2193	PR00011	Type III EGF-like signature IV	PR00011D 12.12 4.57e-12 39-57 IPB001774C 18.25 2.15e-10 68-110
2193	PR00010	Type II EGF-like signature III	PR00010C 6.98 3.90e-10 113-123 PR00011B 13.08 7.88e-10 39-57
2193	IPB000561	EGF-like domain	IPB000561 4.89 9.25e-10 46-54
2193	IPB001886	Laminin N-terminal (Domain VI)	IPB001886E 10.90 9.67e-10 44-60
2193	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 6.21e-09 108-123 PR00011A 14.05 6.88e-09 39-57
2193	IPB000034	Laminin B	IPB000034A 22.21 9.00e-09 96-131
2193	IPB001762	Disintegrin	IPB001762A 23.93 9.65e-09 126-166
2195	IPB000467	D111/G-patch domain	IPB000467 8.65 1.00e-08 329-339
2197	IPB002467	"Methionine aminopeptidase, subfamily 1"	IPB002467C 17.56 2.29e-30 184-212 IPB002467B 12.68 2.50e-23 158-179
2197	PR00599	Methionine aminopeptidase-1 signature II	PR00599B 10.21 8.00e-17 188-204 IPB002467D 14.78 5.50e-15 257-282 PR00599A 11.84 9.63e-14 166-179 IPB002467F 18.38 1.58e-12 315-345 IPB002467E 11.05 7.75e-12 290-302 PR00599D 14.43 5.03e-10 288-300 IPB002467A 15.75 2.87e-09 130-147
2197	IPB001131	Proline dipeptidase	IPB001131D 11.56 5.18e-09 290-303 IPB001131B 18.96 8.10e-09 188-209
2198	IPB002889	WSC domain	IPB002889B 11.76 1.88e-12 366-412 IPB002889B 11.76 3.54e-11 365-411 IPB002889B 11.76 4.96e-10 367-413 IPB002889B 11.76 6.84e-10 363-409 IPB002889B 11.76 7.13e-10 362-408 IPB002889B 11.76 4.19e-09 357-403
2198	IPB003351	Dishevelled specific domain	IPB003351C 13.82 4.49e-09 372-411 IPB002889B 11.76 4.56e-09 353-399 IPB002889B 11.76 7.00e-09 355-401 IPB002889C 9.89 8.52e-09 367-388
2199	PR00918	Calicivirus non-structural polyprotein family signature I	PR00918A 13.81 5.85e-11 192-212
2199	PR00364	Disease resistance protein signature I	PR00364A 8.29 4.71e-09 197-212
2199	PR01102	5-hydroxytryptamine 6 receptor signature XIII	PR01102M 11.13 6.71e-09 1013-1035
2199	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 7.71e-09 1021-1035
2200	IPB001478	PDZ domain (also known as DHR or GLGF)	IPB001478A 11.55 5.09e-09 61-71 IPB001478B 6.12 1.00e-08 79-88
2202	PR01286	Orphan nuclear receptor NOR1 signature V	PR01286E 5.27 9.26e-09 322-343
2203	IPB000998	MAM domain	IPB000998D 18.66 1.96e-15 546-569
2203	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.77e-15 253-272
2203	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 2.89e-14 126-141
2203	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 5.00e-14 208-219 IPB000152 8.86 1.00e-13 253-268 IPB000152 8.86 1.82e-13 208-223 IPB001881B 12.28 4.75e-13 126-137
2203	IPB001774	Delta serrate ligand	IPB001774C 18.25 9.13e-13 88-130 IPB000998B 17.20 1.00e-12 428-440
2203	PR00020	MAM domain signature I	PR00020A 20.48 2.88e-11 426-444 IPB000998C 18.63 5.30e-11 483-498 IPB001881B 12.28 8.58e-11 253-264

TABLE 3B

2203	PR00907	Thrombomodulin signature II	PR00907B 11.50 2.44e-10 160-176
2203	IPB000561	EGF-like domain	IPB000561 4.89 3.25e-10 97-105
2203	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 5.35e-10 258-268 IPB000033B 7.05 5.97e-09 213-223
2203	IPB000167	Dehydrin	IPB000167A 8.58 7.14e-09 340-367
2203	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 9.79e-09 175-195
2204	IPB000998	MAM domain	IPB000998D 18.66 1.96e-15 546-569
2204	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.77e-15 253-272
2204	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 2.89e-14 126-141
2204	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 5.00e-14 208-219 IPB000152 8.86 1.00e-13 253-268 IPB000152 8.86 1.82e-13 208-223 IPB001881B 12.28 4.75e-13 126-137
2204	IPB001774	Delta serrate ligand	IPB001774C 18.25 9.13e-13 88-130 IPB000998B 17.20 1.00e-12 428-440
2204	PR00020	MAM domain signature I	PR00020A 20.48 2.88e-11 426-444 IPB000998C 18.63 5.30e-11 483-498 IPB001881B 12.28 8.58e-11 253-264
2204	PR00907	Thrombomodulin signature II	PR00907B 11.50 2.44e-10 160-176
2204	IPB000561	EGF-like domain	IPB000561 4.89 3.25e-10 97-105
2204	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 5.35e-10 258-268 IPB000033B 7.05 5.97e-09 213-223
2204	IPB000167	Dehydrin	IPB000167A 8.58 7.14e-09 340-367
2204	IPB003367	Thrombospondin type 3 repeat	IPB003367A 11.78 9.79e-09 175-195
2205	IPB002893	MYND zinc finger (ZnF) domain	IPB002893 16.28 4.52e-17 663-681
2205	IPB001664	Intermediate filament proteins	IPB001664B 17.44 6.20e-09 569-608
2205	IPB002889	WSC domain	IPB002889B 11.76 6.34e-09 488-534 IPB002889C 9.89 8.12e-09 437-458 IPB002889B 11.76 9.91e-09 419-465
2206	IPB002893	MYND zinc finger (ZnF) domain	IPB002893 16.28 4.52e-17 663-681
2206	IPB001664	Intermediate filament proteins	IPB001664B 17.44 6.20e-09 569-608
2206	IPB002889	WSC domain	IPB002889B 11.76 6.34e-09 488-534 IPB002889C 9.89 8.12e-09 437-458 IPB002889B 11.76 9.91e-09 419-465
2207	IPB002893	MYND zinc finger (ZnF) domain	IPB002893 16.28 4.52e-17 663-681
2207	IPB001664	Intermediate filament proteins	IPB001664B 17.44 6.20e-09 569-608
2207	IPB002889	WSC domain	IPB002889B 11.76 6.34e-09 488-534 IPB002889C 9.89 8.12e-09 437-458 IPB002889B 11.76 9.91e-09 419-465
2208	IPB002893	MYND zinc finger (ZnF) domain	IPB002893 16.28 4.52e-17 663-681
2208	IPB001664	Intermediate filament proteins	IPB001664B 17.44 6.20e-09 569-608
2208	IPB002889	WSC domain	IPB002889B 11.76 6.34e-09 488-534 IPB002889C 9.89 8.12e-09 437-458 IPB002889B 11.76 9.91e-09 419-465
2210	PR00918	Calicivirus non-structural polyprotein family signature I	PR00918A 13.81 5.85e-11 88-108
2210	PR00364	Disease resistance protein signature I	PR00364A 8.29 4.71e-09 93-108
2211	IPB001762	Disintegrin	IPB001762A 23.93 4.33e-23 19-59
2211	PR00289	Disintegrin signature I	PR00289A 14.29 1.16e-14 35-54 IPB001762B 10.06 3.40e-12 66-76
2211	IPB001774	Delta serrate ligand	IPB001774C 18.25 5.31e-10 238-280 PR00289B 11.74 3.80e-09 64-76
2211	IPB003306	WIF domain	IPB003306E 25.51 7.40e-09 215-260
2212	IPB000159	RA domain	IPB000159A 11.28 7.60e-10 115-124
2212	IPB001359	Synapsin	IPB001359H 22.58 5.89e-09 108-158
2213	PR00308	Type I antifreeze protein signature III	PR00308C 2.79 1.00e-11 729-738

TABLE 3B

2213	IPB000906	ZU5 domain	IPB000906E 22.11 5.55e-11 256-296
2213	PR01415	Ankyrin repeat signature I	PR01415A 12.73 6.46e-11 259-271 IPB000906D 23.89 6.59e-11 324-378 PR01415A 12.73 7.11e-11 192-204 PR01415A 12.73 7.43e-11 160-172 PR00308B 3.38 9.53e-11 729-740 PR00308A 3.72 5.19e-10 726-740 IPB000906F 35.93 5.85e-10 202-255
2213	PR01511	Kv1.4 voltage-gated K ⁺ channel signature IV	PR01511D 3.91 9.26e-10 727-737 PR01415B 10.23 5.88e-09 271-283 IPB000906G 25.85 6.69e-09 338-386 IPB000906A 22.49 7.84e-09 185-227 PR00308A 3.72 9.11e-09 727-741 PR00308C 2.79 9.64e-09 727-736
2214	IPB000471	"Interferon alpha, beta and delta family"	IPB000471A 27.36 2.86e-34 56-109
2214	PR00266	Interferon alpha and beta subunit signature I	PR00266A 13.41 9.59e-14 78-90
2219	PR00405	HIV Rev interacting protein signature II	PR00405B 10.10 2.93e-17 290-307 PR00405A 18.83 4.89e-14 271-290
2219	PR01415	Ankyrin repeat signature I	PR01415A 12.73 1.32e-11 419-431 PR00405C 18.05 2.55e-09 311-332
2220	PR00405	HIV Rev interacting protein signature II	PR00405B 10.10 2.93e-17 290-307 PR00405A 18.83 4.89e-14 271-290
2220	PR01415	Ankyrin repeat signature I	PR01415A 12.73 1.32e-11 419-431 PR00405C 18.05 2.55e-09 311-332
2221	PR00405	HIV Rev interacting protein signature II	PR00405B 10.10 2.93e-17 290-307 PR00405A 18.83 4.89e-14 271-290
2221	PR01415	Ankyrin repeat signature I	PR01415A 12.73 1.32e-11 419-431 PR00405C 18.05 2.55e-09 311-332
2222	PR00405	HIV Rev interacting protein signature II	PR00405B 10.10 2.93e-17 290-307 PR00405A 18.83 4.89e-14 271-290
2222	PR01415	Ankyrin repeat signature I	PR01415A 12.73 1.32e-11 419-431 PR00405C 18.05 2.55e-09 311-332
2223	IPB002870	Reprolysin family propeptide	IPB002870F 18.81 2.35e-19 59-83 IPB002870E 11.90 3.37e-16 23-35
2223	IPB000130	"Neutral zinc metallopeptidases, zinc-binding region"	IPB000130 5.86 1.86e-09 21-31
2223	PR00480	Astacin family signature II	PR00480B 14.35 3.45e-09 16-34
2224	IPB000329	Uteroglobin family	IPB000329A 11.99 3.57e-10 1-16
2224	PR00486	Uteroglobin signature I	PR00486A 6.53 9.03e-09 2-16
2225	IPB001073	Complement C1q protein	IPB001073A 22.14 6.55e-13 67-101
2228	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.09e-11 11-48
2229	IPB001759	Pentaxin family	IPB001759D 18.25 4.67e-33 471-509
2229	PR00895	Pentaxin signature V	PR00895E 12.84 4.19e-18 479-498 PR00895D 14.46 2.38e-17 459-478 PR00895C 12.82 3.18e-17 432-450 IPB001759C 13.49 4.30e-17 432-450 IPB001759A 29.51 1.82e-14 175-209 PR00895A 14.28 8.83e-13 366-380 IPB001759E 18.14 5.34e-11 521-535 PR00895F 15.89 9.50e-11 498-512
2229	IPB002751	Cobalamin synthesis CBIM	IPB002751C 15.32 1.00e-08 50-79
2235	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 8.71e-11 73-98
2239	IPB000917	Sulfatase	IPB000917B 9.25 6.40e-13 103-113 IPB000917A 9.52 5.26e-10 59-70

TABLE 3B

2240	IPB000834	"Zinc carboxypeptidases, carboxypeptidase A metalloprotease (M14) family"	IPB000834B 13.51 2.50e-17 37-51
2240	PR00765	Carboxypeptidase A metalloprotease (M14) family signature II	PR00765B 14.48 1.39e-15 33-47 IPB000834C 17.20 2.80e-15 106-122 IPB000834D 18.95 4.72e-12 133-159 PR00765C 10.88 1.82e-10 113-121
2241	IPB000834	"Zinc carboxypeptidases, carboxypeptidase A metalloprotease (M14) family"	IPB000834B 13.51 2.50e-17 37-51
2241	PR00765	Carboxypeptidase A metalloprotease (M14) family signature II	PR00765B 14.48 1.39e-15 33-47 IPB000834C 17.20 2.80e-15 106-122 IPB000834D 18.95 4.72e-12 133-159 PR00765C 10.88 1.82e-10 113-121
2242	IPB002871	NifU-like N terminal domain	IPB002871C 16.51 1.60e-33 81-113 IPB002871D 14.11 6.87e-21 131-153 IPB002871A 14.39 2.17e-17 35-50 IPB002871B 12.43 6.79e-14 62-74
2244	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 8.29e-11 97-122
2244	PR00048	C2H2-type zinc finger signature II	PR00048B 5.52 9.50e-09 110-119
2245	IPB003527	MAP kinase	IPB003527D 21.53 5.58e-23 214-255 IPB003527G 17.26 8.24e-22 314-351 IPB003527C 14.70 3.05e-19 153-201
2245	IPB001245	Tyrosine kinase catalytic domain	IPB001245A 22.45 5.50e-17 161-201
2245	IPB000959	POLO box duplicated region	IPB000959B 15.68 7.19e-17 145-185 IPB001245B 21.68 1.39e-15 221-259
2245	IPB001772	Kinase associated domain 1	IPB001772C 20.66 3.92e-14 156-186
2245	IPB000095	PAK-box /P21-Rho-binding	IPB000095C 13.36 7.91e-13 75-111 IPB003527A 17.00 6.14e-12 55-80
2245	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861G 13.73 7.44e-12 223-272
2245	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 5.91e-11 217-258 IPB003527B 11.51 9.15e-11 127-145
2245	PR00109	Tyrosine kinase catalytic domain signature II	PR00109B 11.07 9.10e-10 168-186 IPB000961C 15.48 8.83e-09 168-202
2248	IPB001073	Complement C1q protein	IPB001073B 20.88 7.26e-29 42-76
2248	PR00007	Complement C1Q domain signature I	PR00007A 20.64 6.54e-20 35-61 PR00007C 16.13 2.62e-15 110-131 IPB001073C 13.07 1.87e-14 110-129 PR00007B 15.63 3.13e-14 62-81
2250	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.24e-10 325-362
2251	IPB001895	Guanine-nucleotide dissociation stimulators CDC25 family	IPB001895C 20.83 7.84e-30 1097-1132 IPB001895D 18.68 1.00e-20 1194-1217
2251	IPB001331	Guanine-nucleotide dissociation stimulators CDC24 family	IPB001331C 16.09 1.00e-18 397-422 IPB001895B 16.80 3.10e-15 1025-1045 IPB001331B 19.33 7.00e-09 346-361
2253	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 3.91e-09 202-226
2253	PR00169	Potassium channel signature I	PR00169A 17.48 5.50e-09 68-87
2253	IPB002360	Involucrin	IPB002360C 15.36 9.10e-09 198-239
2253	PR01083	Lymphocyte-specific protein signature I	PR01083A 8.60 9.61e-09 214-237
2258	IPB000433	ZZ Zinc finger	IPB000433 14.10 8.20e-18 23-39
2258	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.86e-10 82-107
2261	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 5.91e-11 889-913 IPB000135D 2.13 7.44e-11 897-921

451

TABLE 3B

			IPB000135D 2.13 7.85e-11 899-923 IPB000135D 2.13 3.05e-10 895-919 IPB000135D 2.13 5.11e-10 893-917 IPB000135D 2.13 8.14e-10 900-924 IPB000135D 2.13 2.27e-09 888-912 IPB000135D 2.13 2.27e-09 894-918 IPB000135D 2.13 2.36e-09 892-916
2261	PR00806	Vinculin signature IV	PR00806D 11.95 3.78e-09 577-592 IPB000135D 2.13 3.91e-09 886-910 IPB000135D 2.13 4.45e-09 901-925 IPB000135D 2.13 6.36e-09 896-920 IPB000135D 2.13 7.00e-09 891-915 IPB000135D 2.13 7.18e-09 898-922 IPB000135D 2.13 9.27e-09 932-956
2262	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 6.43e-17 577-601 IPB000135D 2.13 9.71e-17 576-600 IPB000135D 2.13 4.90e-16 580-604 IPB000135D 2.13 8.66e-16 578-602 IPB000135D 2.13 1.13e-15 581-605 IPB000135D 2.13 7.30e-15 579-603 IPB000135D 2.13 7.45e-14 582-606 IPB000135D 2.13 3.08e-13 575-599 IPB000135D 2.13 8.50e-13 584-608 IPB000135D 2.13 8.62e-13 583-607 IPB000135D 2.13 9.08e-13 571-595 IPB000135D 2.13 9.88e-13 586-610 IPB000135D 2.13 1.65e-12 574-598 IPB000135D 2.13 4.36e-12 572-596 IPB000135D 2.13 8.70e-12 585-609 IPB000135D 2.13 8.36e-11 587-611 IPB000135D 2.13 8.67e-11 573-597 IPB000135D 2.13 4.42e-10 567-591 IPB000135D 2.13 3.27e-09 570-594
2262	IPB000637	HMG-I and HMG-Y DNA-binding domain (A+T-hook)	IPB000637B 14.21 4.27e-09 576-594 IPB000135D 2.13 4.45e-09 569-593 IPB000637B 14.21 5.09e-09 585-603
2262	IPB003403	Herpesvirus immediate early protein	IPB003403E 17.25 5.45e-09 577-604 IPB000135D 2.13 7.18e-09 568-592
2262	IPB001422	Neuromodulin (GAP-43)	IPB001422C 16.82 8.54e-09 575-610
2262	IPB001580	Calreticulin family	IPB001580F 2.93 9.10e-09 590-599
2265	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.71e-12 148-185 IPB003006B 20.23 9.14e-12 441-478 IPB003006B 20.23 1.00e-11 248-285
2265	PR01536	Interleukin-1 receptor type I and type II family signature III	PR01536C 19.92 9.23e-11 547-570 IPB003006B 20.23 6.40e-10 54-91 IPB003006B 20.23 9.64e-10 540-577 IPB003006B 20.23 8.62e-09 346-383 PR01536C 19.92 9.19e-09 155-178
2266	IPB000967	Zinc finger NF-X1 type	IPB000967D 10.42 6.89e-09 716-751
2269	IPB002048	EF-hand family	IPB002048 7.91 2.29e-11 178-190
2269	PR00450	Recoverin family signature III	PR00450C 11.99 1.58e-09 64-85 IPB002048 7.91 8.58e-09 105-117
2270	IPB003846	Uncharacterized protein family UPF0061	IPB003846E 18.41 1.00e-40 132-170 IPB003846E 18.41 1.00e-40 511-549 IPB003846F 24.67 9.36e-31 171-206 IPB003846F 24.67 9.36e-31 550-585 IPB003846C 15.01 4.05e-28 8-51 IPB003846G 13.31 5.09e-09 264-274

452
TABLE 3B

			IPB003846G 13.31 5.09e-09 643-653
2271	IPB003846	Uncharacterized protein family UPF0061	IPB003846E 18.41 1.00e-40 132-170 IPB003846E 18.41 1.00e-40 511-549 IPB003846F 24.67 9.36e-31 171-206 IPB003846F 24.67 9.36e-31 550-585 IPB003846C 15.01 4.05e-28 8-51 IPB003846G 13.31 5.09e-09 264-274 IPB003846G 13.31 5.09e-09 643-653
2272	PR00237	Rhodopsin-like GPCR superfamily signature VI	PR00237F 14.34 1.67e-13 51-75 PR00237G 19.23 4.00e-13 89-115
2272	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276B 4.97 6.62e-13 1-12 IPB000276D 9.40 4.52e-10 99-115
2273	PR00019	Leucine-rich repeat signature I	PR00019A 11.72 2.80e-13 89-102 PR00019B 11.42 6.33e-10 86-99
2274	IPB000873	AMP-dependent synthetase and ligase	IPB000873A 11.08 6.06e-14 26-41
2275	IPB000595	Cyclic nucleotide-binding domain	IPB000595B 15.72 6.40e-11 136-159
2276	IPB000595	Cyclic nucleotide-binding domain	IPB000595B 15.72 6.40e-11 136-159
2281	IPB003452	Stem cell factor	IPB003452C 13.68 8.56e-37 207-240
2281	IPB000808	Mrp family	IPB000808A 23.51 1.11e-12 16-60
2281	IPB003348	Anion-transporting ATPase	IPB003348A 20.06 6.60e-11 21-58
2282	PR00205	Cadherin signature VI	PR00205F 19.57 3.37e-17 55-81 PR00205B 20.09 6.67e-16 113-142 PR00205F 19.57 6.70e-13 166-192 PR00205E 10.82 2.17e-10 111-124
2282	IPB002126	Cadherin domain	IPB002126A 14.68 6.09e-10 170-186 PR00205A 17.38 3.12e-09 159-178
2283	PR00205	Cadherin signature VI	PR00205F 19.57 3.37e-17 55-81 PR00205B 20.09 6.67e-16 113-142 PR00205F 19.57 6.70e-13 166-192 PR00205E 10.82 2.17e-10 111-124
2283	IPB002126	Cadherin domain	IPB002126A 14.68 6.09e-10 170-186 PR00205A 17.38 3.12e-09 159-178
2286	IPB002027	Amino acid permease	IPB002027D 22.00 4.13e-25 248-287 IPB002027C 19.67 2.74e-22 167-205 IPB002027B 12.67 7.97e-12 103-122
2287	IPB000559	Formate-tetrahydrofolate ligase	IPB000559C 13.05 1.00e-40 395-444 IPB000559F 12.78 1.00e-40 595-645 IPB000559G 15.54 1.00e-40 649-697 IPB000559D 22.27 4.33e-37 496-536 IPB000559E 17.08 7.39e-36 537-578 IPB000559K 15.77 8.96e-35 875-910 IPB000559B 12.60 2.88e-32 355-383 IPB000559J 17.25 5.94e-32 842-874 IPB000559H 20.31 2.72e-26 712-752 IPB000559A 24.17 6.11e-25 310-354 IPB000559I 15.05 6.35e-18 798-822
2287	PR00085	Tetrahydrofolate dehydrogenase/cyclohydrolase family signature III	PR00085C 13.81 5.70e-14 112-133 PR00085B 16.65 1.23e-09 79-106
2287	IPB000672	Tetrahydrofolate dehydrogenase/cyclohydrolase	IPB000672C 28.03 6.83e-09 153-200
2288	IPB000560	Histidine acid phosphatase	IPB000560 17.02 7.86e-11 391-413
2290	PR00390	Phospholipase C signature I	PR00390A 14.24 6.34e-20 2-20
2292	PR00245	Olfactory receptor signature III	PR00245C 14.65 5.26e-17 183-199 PR00245E 8.96 2.73e-13 290-301 PR00245B 13.73 1.39e-12 136-148

TABLE 3B

			PR00245D 9.34 8.33e-11 243-252
2292	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 1.47e-10 125-136 PR00245A 10.98 8.80e-10 99-110 IPB000276D 9.40 9.61e-10 289-305
2292	PR00896	Vasopressin receptor signature II	PR00896B 9.36 5.50e-09 62-73
2292	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.70e-09 58-70
2292	PR00237	Rhodopsin-like GPCR superfamily signature II	PR00237B 12.45 7.16e-09 66-87 PR00237E 13.03 8.20e-09 206-229
2292	IPB003211	AmiS/UreI family transporter	IPB003211A 15.05 9.43e-09 35-74
2293	IPB003367	Thrombospondin type 3 repeat	IPB003367E 16.82 1.00e-40 35-82 IPB003367F 16.21 1.00e-40 93-142 IPB003367G 17.08 1.00e-40 143-184 IPB003367H 15.25 1.00e-40 185-217 IPB003367J 18.60 1.00e-40 247-288 IPB003367L 21.71 1.00e-40 313-364 IPB003367I 12.15 3.14e-37 218-246 IPB003367K 16.35 9.10e-30 289-312 IPB003367F 16.21 5.83e-21 53-102 IPB003367C 20.73 1.54e-19 38-88 IPB003367D 18.41 9.44e-19 53-95 IPB003367D 18.41 5.55e-17 15-57 IPB003367D 18.41 1.48e-14 93-135 IPB003367F 16.21 2.74e-14 15-64 IPB003367C 20.73 9.27e-13 78-128 IPB003367E 16.82 2.82e-12 12-59 IPB003367E 16.82 4.98e-12 75-122 IPB003367C 20.73 5.96e-11 23-73 IPB003367C 20.73 2.38e-10 101-151 IPB003367C 20.73 6.35e-10 61-111 IPB003367E 16.82 8.88e-10 73-120
2294	IPB001978	Troponin	IPB001978A 18.18 8.89e-09 102-137
2295	IPB000109	PTR peptide transporters (PTR2)	IPB000109D 25.09 6.67e-32 434-481 IPB000109B 29.23 4.18e-23 46-98 IPB000109A 10.85 3.79e-15 23-41 IPB000109C 8.21 7.00e-14 174-186
2295	PR01471	Histamine H3 receptor signature II	PR01471B 12.38 9.63e-09 3-21
2297	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 9.18e-21 113-138 IPB000822 14.67 9.31e-18 29-54 IPB000822 14.67 9.31e-18 141-166 IPB000822 14.67 5.20e-16 57-82 IPB000822 14.67 5.20e-16 85-110
2297	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 4.46e-14 138-151 IPB000822 14.67 1.50e-13 1-26 PR00048A 9.94 5.76e-12 110-123 PR00048A 9.94 1.00e-11 26-39
2297	IPB001275	DM DNA binding domain	IPB001275 19.17 4.21e-11 17-56 PR00048A 9.94 4.79e-11 54-67 IPB001275 19.17 2.22e-10 73-112 PR00048B 5.52 5.50e-10 126-135 IPB001275 19.17 9.15e-10 45-84 PR00048A 9.94 1.38e-09 82-95
2297	IPB001222	TFIIS zinc ribbon domain	IPB001222 24.63 5.69e-09 1-37 IPB001222 24.63 9.49e-09 29-65
2299	IPB003137	Protease associated (PA) domain	IPB003137 22.40 2.50e-19 188-218
2303	IPB000433	ZZ Zinc finger	IPB000433 14.10 8.20e-18 23-39
2303	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.86e-10 82-107

TABLE 3B

2306	IPB001039	"Major histocompatibility complex protein, Class I"	IPB001039A 17.17 1.00e-40 22-75 IPB001039B 27.55 1.00e-40 103-154 IPB001039C 19.82 1.00e-40 184-237 IPB001039D 16.49 1.00e-40 262-316
2306	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.60e-29 268-305 IPB003006A 17.51 6.14e-20 231-253
2306	IPB000353	"Class II histocompatibility antigen, beta chain, beta-1 domain"	IPB000353B 19.16 9.87e-14 210-259
2306	IPB003363	Glycoprotein GG/GX	IPB003363E 13.35 2.94e-11 315-347 IPB000353C 20.11 4.68e-10 261-315
2312	IPB001359	Synapsin	IPB001359H 22.58 5.54e-09 98-148
2312	IPB003403	Herpesvirus immediate early protein	IPB003403A 21.25 6.18e-09 130-152
2313	PR01382	Claudin-9 signature IV	PR01382D 12.38 1.11e-16 205-217
2313	IPB000729	PMP-22/EMP/MP20 family	IPB000729D 18.96 2.96e-16 164-191 IPB000729C 37.83 7.91e-16 84-136 PR01382A 12.00 1.17e-15 41-51
2313	PR01077	Claudin signature III	PR01077C 13.60 1.47e-14 67-77 PR01382C 5.67 5.14e-13 194-203 PR01382B 7.06 1.12e-12 95-104 PR01077B 14.12 1.00e-10 53-59 PR01077D 11.20 4.00e-10 150-156 PR01077A 9.72 8.16e-09 25-34
2317	IPB001245	Tyrosine kinase catalytic domain	IPB001245A 22.45 7.60e-28 129-169
2317	IPB001772	Kinase associated domain 1	IPB001772C 20.66 9.25e-24 124-154
2317	IPB000961	Protein kinase C-terminal domain	IPB000961C 15.48 2.13e-22 136-170 IPB001772D 21.67 4.55e-17 196-235
2317	IPB000959	POLO box duplicated region	IPB000959B 15.68 8.60e-17 113-153
2317	IPB000095	PAK-box /P21-Rho-binding	IPB000095E 17.62 9.03e-17 137-182
2317	IPB003527	MAP kinase	IPB003527C 14.70 1.95e-16 121-169
2317	IPB000861	PKN/rhophilin/rhotekin rho-binding repeat	IPB000861F 16.50 1.55e-15 130-184
2317	IPB000494	"Epidermal growth-factor receptor (EGFR), L domain"	IPB000494C 24.40 7.35e-14 123-169 IPB000959D 27.01 5.95e-13 236-288 IPB000961D 21.23 7.19e-13 185-226 IPB001245B 21.68 8.96e-13 189-227 IPB001772E 24.88 8.96e-12 243-282 IPB003527A 17.00 7.85e-11 28-53 IPB001772A 13.64 2.29e-10 19-50 IPB003527G 17.26 1.30e-09 255-292
2317	PR00109	Tyrosine kinase catalytic domain signature II	PR00109B 11.07 4.23e-09 136-154 IPB003527D 21.53 4.60e-09 182-223
2318	PR01254	Prostaglandin D synthase signature I	PR01254A 12.32 3.37e-29 51-74 PR01254D 13.80 7.97e-27 129-152 PR01254C 10.60 4.68e-22 94-112 PR01254F 10.08 7.58e-21 182-200 PR01254E 14.07 1.00e-18 165-179
2318	PR00179	Lipocalin signature II	PR00179B 7.67 5.26e-13 140-152 PR00179C 17.26 3.84e-12 168-183 PR01254B 12.05 9.04e-12 77-87
2318	PR01275	Neutrophil gelatinase lipocalin signature V	PR01275E 6.38 1.72e-10 135-153 PR00179A 13.97 3.25e-10 57-69
2318	PR01215	Alpha-1-microglobulin signature IV	PR01215D 12.88 9.78e-10 131-150
2318	IPB000566	Lipocalin and cytosolic fatty-acid binding protein	IPB000566B 8.91 1.47e-09 140-150
2318	PR01174	Retinol binding protein signature VI	PR01174F 11.76 3.96e-09 139-155
2318	PR01273	Invertebrate colouration protein signature IV	PR01273D 11.48 4.41e-09 140-154 PR01275B 9.02 8.57e-09 59-69

455

TABLE 3B

2320	IPB001464	Annexin family	IPB001464D 25.42 1.00e-40 177-231 IPB001464B 28.31 1.90e-36 47-99 IPB001464C 24.68 6.40e-30 110-149
2320	PR00196	Annexin family signature IV	PR00196D 21.41 3.81e-22 115-141 PR00196C 9.01 9.67e-22 32-53 PR00196E 9.70 5.22e-21 195-215
2320	PR00201	Annexin type V signature VII	PR00201G 12.46 1.63e-20 195-221
2320	PR00199	Annexin type III signature VI	PR00199F 15.67 5.10e-18 115-141 IPB001464B 28.31 3.86e-17 131-183 PR00196C 9.01 5.70e-17 191-212
2320	PR00200	Annexin type IV signature VII	PR00200G 9.20 7.67e-17 195-221 IPB001464D 25.42 8.71e-17 18-72 PR00199D 4.74 9.87e-17 191-212 PR00199G 9.85 4.45e-16 196-221 PR00196B 11.03 9.31e-16 5-21
2320	PR00197	Annexin type I signature IV	PR00197D 7.59 1.73e-15 32-53 PR00199D 4.74 2.17e-15 32-53 IPB001464A 31.17 3.83e-15 47-101
2320	PR00198	Annexin type II signature IV	PR00198D 7.41 3.89e-15 32-53 PR00197F 9.40 6.80e-15 195-215 PR00200E 8.88 9.02e-15 32-53
2320	PR00202	Annexin type VI signature VII	PR00202G 8.03 9.04e-15 195-221 PR00197D 7.59 1.00e-14 191-212 IPB001464A 31.17 1.85e-14 131-185 PR00198D 7.41 2.38e-14 191-212 PR00198G 7.70 3.44e-13 195-215 PR00201D 8.61 3.51e-13 32-53 PR00200F 14.58 3.53e-13 115-141 P
2321	IPB000175	Sodium:neurotransmitter symporter family	IPB000175C 15.09 1.00e-40 56-107 IPB000175D 23.45 1.00e-40 122-174 IPB000175F 25.63 4.50e-38 310-349 IPB000175E 21.88 5.95e-35 215-254
2321	PR00176	Sodium/chloride neurotransmitter symporter signature V	PR00176E 11.14 2.00e-24 165-185 PR00176G 13.12 3.77e-22 301-321
2321	PR01195	GAT-1 GABA neurotransmitter transporter signature II	PR01195B 13.58 6.60e-22 38-55 PR01195D 9.00 3.75e-21 426-443 PR00176F 11.11 1.36e-19 219-238 IPB000175G 16.18 5.13e-19 371-393 PR00176D 8.96 6.48e-18 83-100 PR00176H 15.94 7.63e-18 341-361 PR01195C 15.62 1.14e-13 191-200
2323	IPB001863	Glypican	IPB001863A 13.95 5.03e-15 56-71
2323	PR00436	Interleukin-8 signature I	PR00436A 15.20 7.91e-10 1-24
2328	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 59-86
2328	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 72-86 IPB001599K 8.15 1.46e-10 29-40
2329	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 59-86
2329	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 72-86 IPB001599K 8.15 1.46e-10 29-40
2330	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 59-86
2330	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 72-86 IPB001599K 8.15 1.46e-10 29-40
2331	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 59-86
2331	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 72-86 IPB001599K 8.15 1.46e-10 29-40
2332	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 59-86
2332	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 72-86

456

TABLE 3B

			IPB001599K 8.15 1.46e-10 29-40
2334	PR00010	Type II EGF-like signature III	PR00010C 6.98 1.37e-11 7-17
2334	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 5.50e-10 2-17
2334	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 8.26e-10 7-17
2335	IPB000492	Protamine 2 (PRM2)	IPB000492B 5.26 7.16e-09 62-96
2336	PR00014	Fibronectin type III repeat signature IV	PR00014D 15.12 5.74e-10 215-229
2339	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 8.36e-35 39-82 IPB002494C 14.46 6.55e-31 83-126 IPB002494C 14.46 9.46e-26 93-136 IPB002494C 14.46 4.84e-25 49-92 IPB002494C 14.46 8.59e-24 44-87 IPB002494C 14.46 9.38e-23 73-116 IPB002494C 14.46 2.73e-22 98-1
2339	IPB000359	Cystine-knot domain	IPB000359B 19.26 9.57e-13 43-61 IPB000359B 19.26 9.57e-13 87-105 IPB002494A 12.44 1.56e-12 61-94 IPB002494B 10.58 2.50e-12 70-84 IPB002494B 10.58 2.50e-12 114-128 IPB002494C 14.46 5.41e-12 53-96
2339	IPB001271	Mammalian defensin	IPB001271 19.97 7.95e-12 77-105 IPB001271 19.97 9.59e-12 38-66 IPB002494B 10.58 1.28e-11 45-59 IPB002494B 10.58 1.28e-11 89-103 IPB002494A 12.44 4.00e-11 75-108
2339	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 4.10e-11 85-130 IPB001271 19.97 5.13e-11 116-144 IPB000006 13.41 6.80e-11 59-104 IPB000359B 19.26 7.48e-11 122-140 IPB000006 13.41 8.00e-11 89-134 IPB002494A 12.44 8.18e-11 65-98 IPB002494C 14.46 1.61e-10 102-
2339	IPB000967	Zinc finger NF-X1 type	IPB000967E 21.88 1.56e-09 70-110
2339	IPB001762	Disintegrin	IPB001762A 23.93 1.88e-09 58-98 IPB001271 19.97 2.15e-09 117-145 IPB002494A 12.44 2.55e-09 81-114 IPB002494A 12.44 3.13e-09 60-93 IPB002494A 12.44 3.23e-09 47-80 IPB002494A 12.44 3.23e-09 91-124 IPB002494A 12.44 3.23e-09 96-1
2340	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 8.36e-35 39-82 IPB002494C 14.46 6.55e-31 83-126 IPB002494C 14.46 9.46e-26 93-136 IPB002494C 14.46 4.84e-25 49-92 IPB002494C 14.46 8.59e-24 44-87 IPB002494C 14.46 9.38e-23 73-116 IPB002494C 14.46 2.73e-22 98-1
2340	IPB000359	Cystine-knot domain	IPB000359B 19.26 9.57e-13 43-61 IPB000359B 19.26 9.57e-13 87-105 IPB002494A 12.44 1.56e-12 61-94 IPB002494B 10.58 2.50e-12 70-84 IPB002494B 10.58 2.50e-12 114-128 IPB002494C 14.46 5.41e-12 53-96
2340	IPB001271	Mammalian defensin	IPB001271 19.97 7.95e-12 77-105 IPB001271 19.97 9.59e-12 38-66

457
TABLE 3B

			IPB002494B 10.58 1.28e-11 45-59 IPB002494B 10.58 1.28e-11 89-103 IPB002494A 12.44 4.00e-11 75-108
2340	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 4.10e-11 85-130 IPB001271 19.97 5.13e-11 116-144 IPB000006 13.41 6.80e-11 59-104 IPB000359B 19.26 7.48e-11 122-140 IPB000006 13.41 8.00e-11 89-134 IPB002494A 12.44 8.18e-11 65-98 IPB002494C 14.46 1.61e-10 102-
2340	IPB000967	Zinc finger NF-X1 type	IPB000967E 21.88 1.56e-09 70-110
2340	IPB001762	Disintegrin	IPB001762A 23.93 1.88e-09 58-98 IPB001271 19.97 2.15e-09 117-145 IPB002494A 12.44 2.55e-09 81-114 IPB002494A 12.44 3.13e-09 60-93 IPB002494A 12.44 3.23e-09 47-80 IPB002494A 12.44 3.23e-09 91-124 IPB002494A 12.44 3.23e-09 96-1
2341	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 8.36e-35 39-82 IPB002494C 14.46 6.55e-31 83-126 IPB002494C 14.46 9.46e-26 93-136 IPB002494C 14.46 4.84e-25 49-92 IPB002494C 14.46 8.59e-24 44-87 IPB002494C 14.46 9.38e-23 73-116 IPB002494C 14.46 2.73e-22 98-1
2341	IPB000359	Cystine-knot domain	IPB000359B 19.26 9.57e-13 43-61 IPB000359B 19.26 9.57e-13 87-105 IPB002494A 12.44 1.56e-12 61-94 IPB002494B 10.58 2.50e-12 70-84 IPB002494B 10.58 2.50e-12 114-128 IPB002494C 14.46 5.41e-12 53-96
2341	IPB001271	Mammalian defensin	IPB001271 19.97 7.95e-12 77-105 IPB001271 19.97 9.59e-12 38-66 IPB002494B 10.58 1.28e-11 45-59 IPB002494B 10.58 1.28e-11 89-103 IPB002494A 12.44 4.00e-11 75-108
2341	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 4.10e-11 85-130 IPB001271 19.97 5.13e-11 116-144 IPB000006 13.41 6.80e-11 59-104 IPB000359B 19.26 7.48e-11 122-140 IPB000006 13.41 8.00e-11 89-134 IPB002494A 12.44 8.18e-11 65-98 IPB002494C 14.46 1.61e-10 102-
2341	IPB000967	Zinc finger NF-X1 type	IPB000967E 21.88 1.56e-09 70-110
2341	IPB001762	Disintegrin	IPB001762A 23.93 1.88e-09 58-98 IPB001271 19.97 2.15e-09 117-145 IPB002494A 12.44 2.55e-09 81-114 IPB002494A 12.44 3.13e-09 60-93 IPB002494A 12.44 3.23e-09 47-80 IPB002494A 12.44 3.23e-09 91-124 IPB002494A 12.44 3.23e-09 96-1
2342	IPB000734	Lipase	IPB000734 10.25 8.12e-09 224-238
2343	IPB000734	Lipase	IPB000734 10.25 8.12e-09 224-238
2344	PR01223	Bride of sevenless protein signature VI	PR01223F 4.19 9.78e-11 205-229
2344	PR00354	7Fe ferredoxin signature III	PR00354C 6.24 8.06e-09 260-277
2345	IPB001304	C-type lectin domain	IPB001304A 17.98 8.04e-14 90-114

TABLE 3B

2345	PR00356	Type II antifreeze protein signature VII	PR00356G 10.21 8.15e-09 201-214
2346	IPB001304	C-type lectin domain	IPB001304A 17.98 8.04e-14 90-114
2346	PR00356	Type II antifreeze protein signature VII	PR00356G 10.21 8.15e-09 201-214
2347	PR00245	Olfactory receptor signature V	PR00245E 8.96 5.15e-16 341-352 PR00245E 8.96 5.15e-16 659-670 PR00245B 13.73 3.77e-15 187-199 PR00245C 14.65 2.73e-14 234-250 PR00245C 14.65 8.27e-14 552-568 PR00245D 9.34 2.59e-13 294-303 PR00245D 9.34 2.59e-13 612-621 PR00245B 13.73 1.39e-12 505-517
2347	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 7.00e-12 176-187 IPB000276A 11.56 7.00e-12 494-505 PR00245A 10.98 8.77e-12 468-479 PR00245A 10.98 1.72e-11 150-161 IPB000276D 9.40 6.09e-10 340-356
2347	PR00237	Rhodopsin-like GPCR superfamily signature II	PR00237B 12.45 7.55e-10 435-456 IPB000276D 9.40 7.65e-10 658-674 PR00237A 9.81 1.84e-09 402-426
2347	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 2.83e-09 109-121 PR00534A 12.77 2.83e-09 427-439 PR00237C 14.77 3.86e-09 162-184 PR00237B 12.45 6.92e-09 117-138 PR00237A 9.81 8.31e-09 84-108
2348	PR00346	Tissue factor signature VIII	PR00346H 10.74 8.18e-09 76-99
2350	PR00457	Animal haem peroxidase signature VII	PR00457G 14.17 4.48e-14 144-164 PR00457H 14.82 5.85e-13 215-229 PR00457F 14.42 6.32e-12 17-27
2351	PR00457	Animal haem peroxidase signature VII	PR00457G 14.17 4.48e-14 144-164 PR00457H 14.82 5.85e-13 215-229 PR00457F 14.42 6.32e-12 17-27
2354	IPB000623	Shikimate kinase	IPB000623A 19.06 6.27e-09 55-84
2360	IPB001841	RING finger	IPB001841 10.69 1.95e-09 159-168
2372	IPB000421	Coagulation factor 5/8 type C domain (FA58C)	IPB000421B 20.70 1.36e-14 129-149
2373	IPB000421	Coagulation factor 5/8 type C domain (FA58C)	IPB000421B 20.70 1.36e-14 129-149
2375	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 3.45e-17 60-98
2375	IPB003527	MAP kinase	IPB003527D 21.53 4.48e-15 53-94
2375	IPB000959	POLO box duplicated region	IPB000959C 23.49 4.21e-12 35-87
2375	IPB000861	PKN/rhopilin/rhotekin rho-binding repeat	IPB000861G 13.73 5.59e-12 62-111
2375	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 2.26e-11 64-118
2375	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 1.61e-10 56-97
2376	IPB001881	Calcium-binding EGF-like domain	IPB001881A 8.72 2.20e-09 41-50
2376	PR00873	Echinoidea (sea urchin) metallothionein signature IV	PR00873D 8.25 8.11e-09 41-59
2377	PR00402	Tec/Btk domain signature I	PR00402A 20.14 8.15e-15 94-113 PR00402B 12.26 4.69e-13 113-125 PR00402C 13.13 8.03e-12 125-138
2379	IPB003886	Extracellular domain in nidogen	IPB003886D 13.91 8.57e-15 46-65
2379	IPB000152	Aspartic acid and asparagine hydroxylation site	IPB000152 8.86 9.05e-14 1-16 IPB000152 8.86 5.91e-13 46-61
2379	IPB001881	Calcium-binding EGF-like domain	IPB001881B 12.28 9.25e-13 1-12
2379	PR01217	Proline rich extensin signature VII	PR01217G 4.02 4.20e-11 125-150

TABLE 3B

2379	IPB000033	"Low-density lipoprotein (ldl) receptor, YWTD repeat"	IPB000033B 7.05 4.96e-11 51-61 IPB001881B 12.28 1.00e-10 46-57
2379	PR00010	Type II EGF-like signature III	PR00010C 6.98 1.66e-09 51-61
2379	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 3.29e-09 133-147 IPB000033B 7.05 3.84e-09 6-16
2379	IPB000561	EGF-like domain	IPB000561 4.89 6.79e-09 55-63 PR00010C 6.98 7.80e-09 6-16
2379	PR00910	Luteovirus ORF6 protein signature I	PR00910A 2.74 8.71e-09 133-145 PR00910A 2.74 9.46e-09 131-143
2385	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 218-234
2385	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 160-171 PR00245D 9.34 1.53e-13 278-287 PR00245E 8.96 6.81e-12 325-336 PR00245B 13.73 1.00e-10 171-183 IPB000276D 9.40 3.08e-09 324-340
2385	PR00237	Rhodopsin-like GPCR superfamily signature V	PR00237E 13.03 3.83e-09 241-264
2385	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.17e-09 93-105 PR00237C 14.77 5.91e-09 146-168
2385	PR00896	Vasopressin receptor signature II	PR00896B 9.36 7.23e-09 97-108 PR00237G 19.23 1.00e-08 314-340
2386	PR00245	Olfactory receptor signature III	PR00245C 14.65 9.53e-17 218-234
2386	IPB000276	Rhodopsin-like GPCR superfamily	IPB000276A 11.56 9.25e-14 160-171 PR00245D 9.34 1.53e-13 278-287 PR00245E 8.96 6.81e-12 325-336 PR00245B 13.73 1.00e-10 171-183 IPB000276D 9.40 3.08e-09 324-340
2386	PR00237	Rhodopsin-like GPCR superfamily signature V	PR00237E 13.03 3.83e-09 241-264
2386	PR00534	Melanocortin receptor family signature I	PR00534A 12.77 5.17e-09 93-105 PR00237C 14.77 5.91e-09 146-168
2386	PR00896	Vasopressin receptor signature II	PR00896B 9.36 7.23e-09 97-108 PR00237G 19.23 1.00e-08 314-340
2389	PR01360	Interleukin-1 receptor antagonist precursor IL-1RA signature VI	PR01360F 14.44 3.11e-12 145-163 PR01360C 10.33 4.84e-11 86-103
2389	IPB000975	Interleukin-1	IPB000975D 24.45 5.55e-09 80-119 IPB000975E 28.12 9.80e-09 124-163
2389	PR00264	Interleukin-1 precursor family signature I	PR00264A 18.63 1.00e-08 83-103
2390	IPB001664	Intermediate filament proteins	IPB001664B 17.44 9.69e-22 102-141 IPB001664C 11.32 4.38e-18 159-186
2390	PR01248	Type I keratin signature II	PR01248B 8.42 6.37e-15 94-117 PR01248C 10.07 9.23e-14 148-168 PR01248A 8.12 4.31e-11 73-86
2390	PR01177	Metabotropic gamma-aminobutyric acid type B1 receptor signature X	PR01177J 6.10 4.96e-10 11-29
2393	PR01276	Type II keratin signature III	PR01276C 10.16 7.32e-11 67-80 PR01276B 9.79 5.96e-10 20-32
2394	IPB001818	Matrixin	IPB001818C 24.38 7.43e-35 54-99 IPB001818B 26.48 8.15e-25 9-50 IPB001818C 24.38 1.55e-21 96-141
2394	PR00138	Matrixin signature III	PR00138C 20.07 1.78e-16 52-80 PR00138B 14.84 5.21e-10 28-43 PR00138C 20.07 9.18e-10 94-122
2395	IPB001818	Matrixin	IPB001818C 24.38 7.43e-35 54-99 IPB001818B 26.48 8.15e-25 9-50 IPB001818C 24.38 1.55e-21 96-141

TABLE 3B

2395	PR00138	Matrixin signature III	PR00138C 20.07 1.78e-16 52-80 PR00138B 14.84 5.21e-10 28-43 PR00138C 20.07 9.18e-10 94-122
2396	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 2.07e-09 10-24
2396	IPB002000	Lysosome-associated membrane glycoprotein (Lamp)	IPB002000D 5.87 5.25e-09 12-25
2405	IPB000364	Phosphoenolpyruvate carboxykinase (GTP)	IPB000364M 26.08 1.40e-09 623-657
2406	IPB001304	C-type lectin domain	IPB001304A 17.98 6.50e-17 155-179
2412	IPB001559	Phosphotriesterase family	IPB001559F 24.25 1.49e-25 343-377 IPB001559D 19.17 5.00e-20 207-233 IPB001559C 16.25 5.34e-16 172-193 IPB001559E 16.18 5.35e-16 245-263 IPB001559A 10.81 1.23e-11 49-60 IPB001559B 12.98 8.50e-10 153-163
2412	IPB000890	Acetate and butyrate kinase	IPB000890E 8.17 8.66e-09 336-349
2414	PR00049	Wilm's tumour protein signature IV	PR00049D 0.00 9.24e-11 410-424 PR00049D 0.00 2.07e-10 412-426 PR00049D 0.00 2.14e-10 411-425 PR00049D 0.00 2.14e-10 414-428
2414	IPB000996	Clathrin light chain	IPB000996B 20.25 8.98e-10 342-394 PR00049D 0.00 9.43e-10 408-422 PR00049D 0.00 9.71e-10 409-423
2414	PR01217	Proline rich extensin signature II	PR01217B 4.82 7.09e-09 412-428
2414	IPB002999	Tudor domain	IPB002999B 7.50 7.55e-09 412-420
2414	PR01471	Histamine H3 receptor signature V	PR01471E 5.41 8.92e-09 411-426 PR00049D 0.00 8.93e-09 413-427
2415	PR01372	Yersinia virulence determinant YopE protein signature II	PR01372B 7.73 4.87e-09 21-38
2420	IPB003817	Phosphatidylserine decarboxylase	IPB003817D 23.34 8.71e-25 194-220 IPB003817C 10.66 4.00e-15 172-184 IPB003817E 13.21 2.67e-14 283-299 IPB003817A 12.64 4.15e-13 77-91 IPB003817B 13.04 4.00e-09 101-109
2425	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469J 8.97 3.52e-12 17-33
2426	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469J 8.97 3.52e-12 17-33
2427	IPB002469	"Dipeptidyl peptidase IV, N-terminus"	IPB002469J 8.97 3.52e-12 17-33
2429	IPB000906	ZU5 domain	IPB000906A 22.49 6.14e-19 145-187 IPB000906F 35.93 3.09e-16 63-116 IPB000906F 35.93 7.91e-16 96-149
2429	PR01415	Ankyrin repeat signature I	PR01415A 12.73 3.70e-15 252-264 IPB000906A 22.49 1.71e-14 46-88 IPB000906F 35.93 1.00e-12 346-399 IPB000906A 22.49 5.66e-12 112-154 IPB000906G 25.85 9.36e-12 53-101 PR01415A 12.73 1.00e-11 53-65 PR01415A 12.73 2.61e-11 119-13
2430	PR00834	HtrA/DegQ protease family signature III	PR00834C 15.48 7.35e-19 148-172 PR00834D 11.75 7.39e-17 186-203 PR00834B 10.17 3.25e-13 107-127 PR00834E 13.43 6.03e-12 208-225
2430	IPB000126	"Serine proteases, V8 family"	IPB000126B 12.50 6.81e-12 191-207 PR00834A 8.79 1.44e-11 86-98 PR00834F 11.11 1.53e-09 301-313

461
TABLE 3B

			IPB000126A 11.75 9.83e-09 78-93
2432	PR00505	D12 class N6 adenine-specific DNA methyltransferase signature I	PR00505A 15.44 3.67e-12 39-55 PR00505B 11.79 8.88e-12 60-74
2433	PR00179	Lipocalin signature II	PR00179B 7.67 2.35e-09 15-27 PR00179C 17.26 6.70e-09 42-57
2433	PR01174	Retinol binding protein signature VI	PR01174F 11.76 6.82e-09 14-30
2433	PR01254	Prostaglandin D synthase signature V	PR01254E 14.07 8.23e-09 39-53
2434	PR01042	Aspartyl-tRNA synthetase signature II	PR01042B 12.76 4.69e-11 260-273 PR01042A 9.01 9.77e-10 244-256
2434	IPB002106	Aminoacyl-transfer RNA synthetases class-II	IPB002106A 13.35 1.00e-08 196-208
2435	IPB003952	Fumarate reductase / succinate dehydrogenase FAD-binding site	IPB003952D 19.72 4.50e-20 7-35 IPB003952E 9.04 2.46e-16 48-65
2436	IPB001895	Guanine-nucleotide dissociation stimulators CDC25 family	IPB001895C 20.83 8.50e-23 52-87
2437	IPB000958	KH domain	IPB000958 6.84 5.09e-12 173-186 IPB000958 6.84 2.29e-11 89-102
2440	IPB001393	Calsequestrin	IPB001393A 16.72 1.00e-40 66-115 IPB001393B 11.93 1.00e-40 169-222 IPB001393C 16.33 1.00e-40 225-277 IPB001393D 11.26 1.00e-40 320-372
2440	PR00312	Calsequestrin signature V	PR00312E 8.61 7.75e-36 200-229 PR00312I 15.97 5.71e-35 363-391 PR00312H 13.19 2.80e-34 294-321 PR00312J 13.61 6.48e-34 394-422 PR00312D 9.10 7.17e-33 159-188 PR00312B 14.57 4.41e-32 93-122 PR00312C 16.48 5.62e-32 123-152 PR00312G 11.43 1.49e-31 261-288 PR00312F 16.12 1.73e-31 230-259 PR00312A 11.96 7.94e-27 66-89
2442	IPB000353	"Class II histocompatibility antigen, beta chain, beta-1 domain"	IPB000353B 19.16 4.94e-16 139-188
2442	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006A 17.51 8.50e-16 160-182
2442	IPB001003	"MHC Class II, alpha chain, alpha-1 domain"	IPB001003B 14.72 9.90e-10 147-190
2444	PR00021	Small proline-rich protein signature I	PR00021A 3.31 1.35e-19 8-20 PR00021B 5.91 1.00e-14 31-40 PR00021B 5.91 1.00e-13 22-31 PR00021D 4.82 1.39e-13 25-33 PR00021D 4.82 1.39e-13 34-42 PR00021D 4.82 6.87e-13 43-51 PR00021B 5.91 1.92e-11 40-49 PR00021E 7.77 1.23e-10 61-70 PR00021C 5.97 1.25e-10 25-31 PR00021C 5.97 1.25e-10 34-40
2444	PR01217	Proline rich extensin signature IV	PR01217D 4.57 4.94e-10 30-51 PR01217G 4.02 2.42e-09 23-48 PR01217G 4.02 2.42e-09 30-55 PR01217G 4.02 2.58e-09 21-46 PR01217D 4.57 7.89e-09 21-42 PR01217G 4.02 8.89e-09 39-64
2444	IPB000967	Zinc finger NF-X1 type	IPB000967E 21.88 9.44e-09 12-52
2445	PR00205	Cadherin signature VI	PR00205F 19.57 5.15e-21 522-548 PR00205B 20.09 5.50e-21 254-283 PR00205D 12.22 1.39e-15 338-357

462
TABLE 3B

			PR00205B 20.09 2.50e-15 464-493 PR00205D 12.22 6.09e-15 233-252 PR00205G 13.05 8.00e-15 556-573 PR00205A 17.38 2.59e-14 85-104
2445	IPB002126	Cadherin domain	IPB002126B 12.04 4.79e-14 242-259 PR00205B 20.09 5.63e-13 145-174 IPB002126B 12.04 7.43e-13 452-469 PR00205D 12.22 7.60e-13 443-462 PR00205G 13.05 7.75e-13 341-358 PR00205F 19.57 3.38e-12 309-335 PR00205G 13.05 9.10e-12 236-253 PR00205E 10.82 3.37e-11 252-265 PR00205E 10.82 7.16e-11 462-475 PR00205D 12.22 7.59e-11 553-572 PR00205F 19.57 9.05e-11 412-438 IPB002126A 14.68 4.91e-10 206-222 IPB002126A 14.68 5.30e-10 416-432 IPB002126B 12.04 3.25e-09 133-150 PR00205C 13.59 3.25e-09 326-338 IPB002126B 12.04 4.50e-09 347-364 PR00205B 20.09 9.83e-09 581-610 PR00205G 13.05 1.00e-08 446-463
2447	IPB000006	"Vertebrate metallothionein, family 1"	IPB000006 13.41 3.90e-12 29-74 IPB000006 13.41 4.41e-12 36-81 IPB000006 13.41 6.70e-11 32-77
2447	PR01228	Eggshell protein signature III	PR01228C 5.69 1.22e-10 23-38 PR01228C 5.69 1.98e-10 7-22
2447	IPB001271	Mammalian defensin	IPB001271 19.97 3.29e-10 48-76
2447	IPB002494	"Keratin, high sulfur B2 protein"	IPB002494C 14.46 3.36e-10 42-85 IPB001271 19.97 3.47e-10 26-54 IPB002494A 12.44 6.11e-10 67-100
2447	IPB002174	Furin-like cysteine rich region	IPB002174A 30.51 7.32e-10 8-39 PR01228C 5.69 8.05e-10 16-31
2447	IPB003571	Snake toxin	IPB003571B 18.08 8.07e-10 73-96 IPB002494A 12.44 9.08e-10 22-55
2447	PR00858	Crustacean metallothionein signature II	PR00858B 5.93 1.48e-09 37-55 IPB000006 13.41 3.11e-09 33-78
2447	IPB001169	"Integrin beta, C-terminus"	IPB001169K 27.45 3.19e-09 39-81
2447	IPB002919	Trypsin Inhibitor-like cysteine rich domain	IPB002919A 15.56 3.57e-09 49-61 IPB002174A 30.51 4.15e-09 24-55 IPB001271 19.97 4.44e-09 55-83 IPB002494A 12.44 4.97e-09 29-62 PR01228C 5.69 5.03e-09 15-30 PR01228C 5.69 5.03e-09 19-34 IPB002174A 30.51 5.28e-09 16-47
2447	IPB000254	"Cellulose-binding domain, fungal type"	IPB000254 18.11 5.36e-09 25-55 IPB000006 13.41 5.59e-09 39-84 IPB002174A 30.51 5.72e-09 33-64 PR01228C 5.69 5.76e-09 24-39
2447	IPB000867	Insulin-like growth factor-binding protein	IPB000867B 11.44 6.55e-09 2-18 IPB002174A 30.51 6.62e-09 4-35
2447	IPB002867	Cysteine-rich domain (C6HC)	IPB002867D 24.88 7.19e-09 35-66 IPB000006 13.41 7.24e-09 47-92
2447	IPB000967	Zinc finger NF-X1 type	IPB000967D 10.42 7.37e-09 57-92 IPB001169K 27.45 7.81e-09 32-74 IPB000006 13.41 8.07e-09 37-82 IPB002494A 12.44 8.35e-09 26-59

463
TABLE 3B

			IPB000006 13.41 8.44e-09 52-97
2447	PR01117	CLC-6 chloride channel signature I	PR01117A 7.79 9.47e-09 48-60 IPB001271 19.97 9.51e-09 64-92 IPB002174A 30.51 9.77e-09 36-67
2447	IPB002221	WAP-type (Whey Acidic Protein) four-disulfide core domain	IPB002221B 17.12 1.00e-08 45-66
2448	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 4.79e-12 52-77
2448	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 3.05e-10 49-62 IPB000822 14.67 9.14e-10 200-225
2450	PR00946	Mercury scavenger protein signature I	PR00946A 4.14 8.16e-09 6-24
2452	IPB002038	Osteopontin	IPB002038C 22.35 1.00e-40 173-214
2452	PR00216	Osteopontin signature I	PR00216A 11.45 9.71e-34 43-72 IPB002038A 12.23 5.15e-31 42-71 PR00216C 9.12 7.82e-21 95-120 PR00216B 6.70 9.49e-21 79-108 PR00216D 3.16 3.30e-18 142-156 PR00216E 6.95 3.81e-18 174-188 IPB002038B 15.58 4.11e-16 77-121 PR00216D 3.16 3.69e-12 136-150
2452	IPB003403	Herpesvirus immediate early protein	IPB003403E 17.25 9.26e-09 117-144 IPB002038B 15.58 9.58e-09 91-135
2454	IPB001241	DNA topoisomerase II family	IPB001241F 23.94 8.36e-37 475-523
2454	PR01158	Topoisomerase II signature VIII	PR01158H 13.39 5.50e-30 804-826 IPB001241G 14.13 1.00e-29 547-573 PR01158K 14.14 5.24e-27 1023-1049 PR01158G 9.37 5.91e-27 757-780
2454	IPB002205	"DNA gyrase/topoisomerase IV, subunit A"	IPB002205B 14.49 4.79e-24 760-795 IPB001241E 20.94 3.00e-22 371-397 PR01158I 13.95 7.00e-22 834-854 PR01158D 11.94 5.24e-21 565-580
2454	PR00418	DNA topoisomerase II family signature VI	PR00418F 13.13 3.40e-20 546-562 IPB001241A 15.98 6.04e-20 50-71 IPB001241B 10.04 2.71e-19 172-190 PR00418G 12.91 8.94e-19 564-581 IPB001241H 17.27 1.96e-18 808-831
2454	PR00615	CCAAT-binding transcription factor subunit A signature I	PR00615A 17.09 2.93e-18 319-337 PR01158J 13.56 3.45e-18 939-953 IPB002205D 10.13 3.54e-18 867-888 PR00615B 18.03 3.77e-18 707-725 PR00418C 9.38 1.82e-17 176-190 PR00418I 17.21 4.60e-17 626-642 IPB002205A 8.13 9.54e-17 729-747 PR00418A 13.58 7.65e-16 96-111 PR01158C 11.35 1.00e-15 519-532 PR01158E 8.11 2.29e-15 585-596 PR01158F 10.39 4.71e-15 632-644 PR00615C 17.93 8.50e-15 1148-1166 PR00418E 14.82 1.37e-14 473-487 IPB001241D 14.87 1.43e-14 328-341 PR00418B 12.37 2.57e-14 133-146 PR00418D 14.25 2.71e-14 328-341 PR01158A 7.61 4.60e-13 456-466 IPB002205C 11.89 5.09e-12 812-826 PR00418H 10.58 5.91e-12 584-596 IPB001241C 13.37 1.31e-11 230-242
2454	IPB000509	Ribosomal protein L36E	IPB000509B 20.29 7.85e-11 1216-1270

TABLE 3B

			PR01158B 8.30 1.27e-10 471-478
2454	IPB000135	High mobility group proteins HMG1 and HMG2	IPB000135D 2.13 5.64e-09 1362-1386 IPB000135D 2.13 7.45e-09 1363-1387 IPB000135D 2.13 8.09e-09 1364-1388
2454	PR01469	Bacterial carbamate kinase signature V	PR01469E 10.60 8.43e-09 128-146 IPB000135D 2.13 8.73e-09 1360-1384
2457	IPB001073	Complement C1q protein	IPB001073A 22.14 6.55e-13 67-101
2466	IPB000959	POLO box duplicated region	IPB000959D 27.01 9.61e-10 204-256
2473	PR01475	Parkin signature IX	PR01475I 10.01 8.01e-09 96-118
2476	IPB003743	DUF164	IPB003743B 20.16 4.64e-09 88-126
2481	IPB000215	Serpins	IPB000215C 13.90 5.00e-09 435-449
2482	PR01377	Claudin-1 signature IV	PR01377D 6.30 1.00e-19 229-243 PR01377A 7.94 1.00e-16 141-152
2482	IPB000729	PMP-22/EMP/MP20 family	IPB000729D 18.96 5.50e-15 197-224
2482	PR01077	Claudin signature III	PR01077C 13.60 2.53e-12 99-109 PR01377B 13.79 1.12e-11 176-183 PR01377C 14.12 2.44e-11 188-195 PR01077B 14.12 1.00e-10 85-91 IPB000729C 37.83 5.31e-10 116-168 PR01077A 9.72 4.49e-09 57-66
2482	PR01385	Claudin-14 signature I	PR01385A 5.13 5.70e-09 46-62
2483	IPB001919	"Cellulose-binding domain, bacterial type"	IPB001919B 14.22 2.97e-09 188-212
2487	PR01305	Invasion protein B family signature IV	PR01305D 7.82 6.19e-09 266-279
2488	IPB002652	Importin beta binding domain	IPB002652H 25.98 1.00e-40 568-614 IPB002652I 18.58 1.36e-35 647-683
2488	IPB000225	Armadillo repeat	IPB000225E 20.58 8.20e-22 646-668 IPB002652C 21.73 5.88e-14 519-571 IPB000225D 18.99 5.02e-13 535-558 IPB002652F 18.67 9.25e-11 543-575 IPB002652G 22.45 1.36e-09 535-580
2488	IPB003191	Guanylate-binding protein	IPB003191M 10.38 7.64e-09 69-99
2490	IPB001762	Disintegrin	IPB001762A 23.93 4.33e-23 19-59
2490	PR00289	Disintegrin signature I	PR00289A 14.29 1.16e-14 35-54 IPB001762B 10.06 3.40e-12 66-76
2490	IPB001774	Delta serrate ligand	IPB001774C 18.25 5.31e-10 238-280 PR00289B 11.74 3.80e-09 64-76
2490	IPB003306	WIF domain	IPB003306E 25.51 7.40e-09 215-260
2491	IPB001359	Synapsin	IPB001359H 22.58 6.07e-09 96-146
2495	IPB001359	Synapsin	IPB001359H 22.58 6.33e-09 35-85 IPB001359H 22.58 7.73e-09 41-91
2496	IPB001359	Synapsin	IPB001359H 22.58 6.33e-09 35-85 IPB001359H 22.58 7.73e-09 41-91
2497	IPB001359	Synapsin	IPB001359H 22.58 6.33e-09 35-85 IPB001359H 22.58 7.73e-09 41-91
2498	IPB000492	Protamine 2 (PRM2)	IPB000492B 5.26 7.95e-09 230-264
2502	PR01415	Ankyrin repeat signature I	PR01415A 12.73 1.25e-09 187-199
2502	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 9.31e-09 89-126
2504	IPB000492	Protamine 2 (PRM2)	IPB000492B 5.26 1.68e-09 219-253
2504	PR00580	Prostanoid EP1 receptor signature V	PR00580E 8.05 7.11e-09 226-247
2505	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.54e-09 195-232
2506	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 6.54e-09 195-232
2507	PR00456	Ribosomal protein P2 signature V	PR00456E 3.08 9.42e-10 637-651

465
TABLE 3B

2508	PR01481	Neurotensin type 2 receptor signature III	PR01481C 15.05 1.00e-17 176-189
2508	PR01479	Neurotensin receptor signature II	PR01479B 12.40 2.43e-17 89-101 PR01481A 7.58 3.54e-16 1-13 PR01479C 7.31 1.00e-15 102-115 PR01481B 6.68 1.45e-15 14-26 PR01481D 4.62 2.19e-15 190-201 PR01479E 8.74 3.70e-15 240-250 PR01479D 13.10 6.57e-14 229-239 PR01479A 8.89 1.00e-13 29-39
2508	PR00237	Rhodopsin-like GPCR superfamily signature VII	PR00237G 19.23 4.44e-12 249-275
2508	PR00665	Oxytocin receptor signature IV	PR00665D 10.30 1.32e-11 134-150 PR01479F 8.03 5.19e-11 277-287 PR00237C 14.77 4.32e-10 115-137 PR00237A 9.81 7.33e-10 34-58 PR00237D 9.76 7.43e-10 151-172
2508	PR01417	Growth hormone secretagogue receptor type 1 signature IV	PR01417D 12.33 8.13e-10 111-127 PR00237F 14.34 6.05e-09 204-228
2509	IPB001101	Plectin repeat	IPB001101A 10.14 5.40e-14 1-37
2510	IPB001101	Plectin repeat	IPB001101A 10.14 5.40e-14 1-37
2517	IPB001552	Acyl-CoA dehydrogenase	IPB001552E 22.77 2.46e-19 523-563 IPB001552D 24.88 5.35e-19 432-474 IPB001552C 25.04 7.75e-15 378-418 IPB001552B 18.05 3.43e-12 124-146 IPB001552A 11.25 6.90e-10 97-108
2518	IPB001552	Acyl-CoA dehydrogenase	IPB001552E 22.77 2.46e-19 523-563 IPB001552D 24.88 5.35e-19 432-474 IPB001552C 25.04 7.75e-15 378-418 IPB001552B 18.05 3.43e-12 124-146 IPB001552A 11.25 6.90e-10 97-108
2519	IPB002524	Cation efflux family	IPB002524B 23.89 5.20e-17 50-89
2519	IPB003452	Stem cell factor	IPB003452B 19.11 6.63e-09 109-157 IPB002524A 20.13 7.39e-09 8-48
2520	PR00215	Neuromodulin signature III	PR00215C 13.82 7.58e-10 478-498
2520	PR00194	Tropomyosin signature IV	PR00194D 9.54 7.19e-09 357-380
2520	IPB001422	Neuromodulin (GAP-43)	IPB001422A 13.23 7.43e-09 453-497
2521	PR01178	Metabotropic gamma-aminobutyric acid type B2 receptor signature XI	PR01178K 13.44 8.65e-09 179-203
2523	IPB002889	WSC domain	IPB002889B 11.76 4.56e-10 34-80 IPB002889B 11.76 7.84e-09 19-65 IPB002889B 11.76 7.84e-09 27-73 IPB002889B 11.76 1.00e-08 23-69
2529	PR00019	Leucine-rich repeat signature II	PR00019B 11.42 1.33e-10 225-238 PR00019A 11.72 8.33e-10 228-241 PR00019A 11.72 4.00e-09 202-215 PR00019B 11.42 7.82e-09 199-212
2530	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.60e-10 297-334
2530	IPB001000	Glycoside hydrolase family 10	IPB001000H 10.38 7.80e-09 13-26
2531	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 4.60e-10 297-334
2531	IPB001000	Glycoside hydrolase family 10	IPB001000H 10.38 7.80e-09 13-26
2532	IPB003884	Factor I membrane attack complex	IPB003884A 12.20 7.06e-09 56-67
2536	IPB000822	"Zinc finger, C2H2 type"	IPB000822 14.67 7.50e-13 309-334
2536	PR00048	C2H2-type zinc finger signature I	PR00048A 9.94 4.18e-12 306-319 IPB000822 14.67 5.74e-12 281-306

466

TABLE 3B

2536	PR00258	Speract receptor signature I	PR00258A 13.56 2.98e-10 87-103
2536	IPB002867	Cysteine-rich domain (C6HC)	IPB002867C 19.46 9.25e-10 306-323
2540	IPB001522	"Fatty acid desaturase, type 1"	IPB001522F 22.32 1.00e-40 104-158 IPB001522E 20.55 5.85e-36 26-79
2540	PR00075	Fatty acid desaturase family 1 signature VII	PR00075G 10.50 6.62e-20 131-145 PR00075E 11.60 6.46e-18 55-73 PR00075F 14.62 8.81e-16 88-109
2541	IPB000432	"DNA mismatch repair protein MutS family, C-terminal domain"	IPB000432D 18.83 8.92e-39 369-417 IPB000432C 12.07 1.00e-37 329-360 IPB000432F 16.97 3.86e-27 476-507 IPB000432E 8.78 9.00e-13 441-451
2541	IPB002156	RNase H	IPB002156B 11.33 2.20e-11 100-110
2542	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 8.20e-10 33-70
2543	IPB000998	MAM domain	IPB000998C 18.63 1.95e-12 17-32
2543	PR00020	MAM domain signature III	PR00020C 12.01 8.12e-10 16-27 IPB000998D 18.66 9.61e-10 82-105
2544	IPB002350	Kazal-type serine protease inhibitor family	IPB002350 31.78 3.92e-13 46-86
2544	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 1.78e-11 150-187
2545	PR00449	Transforming protein P21 ras signature I	PR00449A 12.48 8.16e-10 86-107
2545	PR00326	GTP1/OBG GTP-binding protein family signature I	PR00326A 8.70 9.13e-10 88-108
2545	IPB000619	Guanylate kinase	IPB000619A 18.08 4.21e-09 88-105
2545	PR00364	Disease resistance protein signature I	PR00364A 8.29 7.14e-09 87-102
2545	PR00094	Adenylate kinase signature I	PR00094A 9.62 9.57e-09 89-102
2545	PR00918	Calicivirus non-structural polyprotein family signature I	PR00918A 13.81 9.69e-09 82-102
2545	IPB000795	GTP-binding elongation factor	IPB000795A 10.67 9.77e-09 87-102
2547	IPB003006	Immunoglobulin and major histocompatibility complex domain	IPB003006B 20.23 3.08e-09 4-41
2548	PR00698	C.elegans Srg family integral membrane protein signature V	PR00698E 14.65 2.76e-09 95-120
2551	IPB001737	Ribosomal RNA adenine dimethylase	IPB001737A 27.11 8.54e-10 135-180
2553	IPB000906	ZU5 domain	IPB000906A 22.49 3.16e-09 38-80
2554	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 6.54e-13 281-319
2554	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 3.97e-11 285-339
2554	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 2.22e-10 277-318 IPB001245A 22.45 3.18e-10 228-268
2555	IPB001245	Tyrosine kinase catalytic domain	IPB001245B 21.68 6.54e-13 281-319
2555	IPB000095	PAK-box /P21-Rho-binding	IPB000095F 16.47 3.97e-11 285-339
2555	IPB000961	Protein kinase C-terminal domain	IPB000961D 21.23 2.22e-10 277-318 IPB001245A 22.45 3.18e-10 228-268
2557	PR01041	Methionyl-tRNA synthetase signature V	PR01041E 16.72 2.69e-17 60-75 PR01041D 11.02 7.43e-13 30-41
2557	IPB001412	Aminoacyl-transfer RNA synthetases class-I	IPB001412B 6.33 8.71e-12 98-108
2558	IPB000353	"Class II histocompatibility antigen, beta chain, beta-1 domain"	IPB000353A 18.51 7.30e-27 41-90
2563	IPB001599	Alpha-2-macroglobulin family	IPB001599L 18.66 4.15e-28 59-86
2563	IPB001134	"Netrin, C-terminus"	IPB001134C 17.82 4.13e-13 72-86 IPB001599K 8.15 1.46e-10 29-40

467

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
686	hormone	Somatotropin hormone family	3.1e-27	103.9	1	9-182
688	hormone	Somatotropin hormone family	4.2e-37	136.7	1	9-176
689	serpin	Serpin (serine protease inhibitor)	1.8e-74	260.8	1	51-397
690	efhand	EF hand	2.7e-08	41.0	2	34-62:70-98
691	Lipase_3	Lipase (class 3)	2.3e-20	81.1	1	366-505
692	PH	PH domain	0.028	21.0	1	36-127
694	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) family	4.2e-51	183.2	1	93-483
695	7tm_1	7 transmembrane receptor (rhodopsin family)	3.3e-21	83.9	1	22-294
696	lectin_c	Lectin C-type domain	5.1e-06	33.3	1	181-286
698	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) family	3.8e-42	153.5	1	40-402
700	myb_DNA-binding	Myb-like DNA-binding domain	9.3e-09	42.5	1	231-278
700	ZZ	Zinc finger, ZZ type	0.021	17.8	1	168-211
702	zf-AN1	AN1-like Zinc finger	0.0034	18.0	2	10-52:103-138
703	CRAL_TRIO	CRAL/TRIO domain	2.5e-41	150.7	1	85-280
703	CRAL_TRIO_N	CRAL/TRIO, N-terminus	5.9e-10	46.5	1	3-71
704	Rhomboid	Rhomboid family	0.019	-10.9	1	152-307
705	GKAP	Guanylate-kinase-associated protein (GKAP) p	7e-292	983.1	1	621-979
706	LBP_BPI_CE TP_C	LBP / BPI / CETP family, C-terminal do	4.6e-06	33.6	1	218-456
707	Glyco_transf_8	Glycosyl transferase family 8	0.0021	-38.4	1	103-368
708	LIM	LIM domain	7.8e-14	59.4	1	12-68
710	Collagen	Collagen triple helix repeat (20 copies)	8e-169	574.2	20	56-114:115-174:187-245:291-349:360-418:423-483:492-550:598-656:684-743:750-808:809-868:869-928:929-988:1032-1090:1096-1154:1155-1214:1217-1277:1278-1337:1341-1400:1417-1476
710	C4	C-terminal tandem repeated domain in type 4	1.5e-148	506.9	2	1489-1596:1597-1711
711	ldl_recept_a	Low-density lipoprotein receptor domain	0	1307.3	32	67-108:112-152:880-920:921-961:962-1001:1002-1041:1042-1081:1088-1127:1130-1170:1171-1212:2545-2586:2587-2625:2626-2664:2676-

468

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						2713:2717- 2755:2756- 2795:2796- 2838:2840- 2879:2880- 2923:2926- 2964:3352- 3391:3392- 3430:3431- 3470:3471- 3510:3511- 3549:3550- 3588:3589- 3626:3629- 3667:3668- 3706:3709- 3749:3750- 3790:3797-3835
711	ldl_recept_b	Low-density lipoprotein receptor repeat	2.4e-239	808.6	34	332-373:375- 417:419-461:605- 646:648-692:694- 742:744-791:1337- 1382:1384- 1425:1427- 1472:1474- 1517:1518- 1558:1655- 1696:1698- 1740:1742- 1780:1782- 1825:1959- 2000:2002- 2043:2045- 2087:2089- 2131:2276- 2315:2318- 2365:2367- 2410:2412- 2453:2454- 2495:3092- 3134:3136- 3177:3179- 3221:3223- 3260:3262- 3303:3970- 4016:4018- 4074:4076- 4118:4120-4163
711	EGF	EGF-like domain	1.8e-28	108.0	36	69-106:157-190:196- 230:512-553:835- 870:1004-1039:1043- 1079:1090- 1125:1173- 1210:1213- 1249:1255- 1289:1568-

469

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						1606:1875- 1911:2184- 2219:2505- 2540:2589- 2623:2635- 2662:2719- 2753:2928- 2962:2967- 3003:3009- 3041:3314- 3350:3513- 3547:3552- 3586:3590- 3624:3669- 3704:3752- 3788:3842- 3879:3885- 3917:4213- 4244:4254- 4285:4290- 4321:4326- 4357:4362- 4393:4398- 4428:4431-4463
712	ldl_recept_a	Low-density lipoprotein receptor domain	4.7e-21	83.4	2	67-108:112-152
714	cadherin	Cadherin domain	0	1168.1	16	47-126:140-241:255- 344:363-466:480- 573:588-680:694- 784:798-884:898- 987:1001-1091:1105- 1201:1215- 1306:1320- 1411:1425- 1520:1526- 1622:1634-1728
715	cadherin	Cadherin domain	0	1177.0	16	47-126:140-241:255- 344:363-466:480- 573:588-680:694- 784:798-884:898- 987:1001-1091:1105- 1201:1215- 1306:1320- 1411:1425- 1520:1526- 1622:1634-1729
716	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	1.2e-07	-81.3	1	132-652
716	Peptidase_S9	Prolyl oligopeptidase family	1.7e-06	35.0	1	656-736
717	zf-C2H2	Zinc finger, C2H2 type	3.6e-71	249.9	10	32-54:60-82:154- 176:182-204:210- 232:238-260:266- 288:294-316:322- 344:350-372
720	ig	Immunoglobulin domain	2.8e-178	605.6	15	68-128:163-223:259- 317:352-410:445-

470

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						503:538-596:629-687:720-780:813-871:904-962:995-1052:1085-1143:1176-1232:1266-1323:1356-1413
720	tsp_1	Thrombospondin type 1 domain	5e-87	302.5	6	1435-1485:1492-1542:1549-1599:1606-1656:1663-1713:1720-1770
720	EGF	EGF-like domain	1.6e-32	121.5	8	2013-2047:2053-2092:2098-2130:2136-2172:2178-2215:2221-2256:2338-2372:2378-2418
721	SPRY	SPRY domain	2.7e-29	110.7	1	289-418
721	SAP	SAP domain	6.9e-09	43.0	1	3-37
722	ABC_tran	ABC transporter	1e-105	364.6	2	510-692:1322-1506
724	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal domain	1.6e-49	178.0	1	50-201
725	EGF	EGF-like domain	1.9e-18	74.7	5	65-91:98-132:138-172:178-217:223-258
725	MAM	MAM domain	1.7e-13	58.3	1	402-546
726	NHL	NHL repeat	5.4e-67	236.0	6	431-458:478-505:525-552:572-599:619-646:666-693
726	Filamin	Filamin/ABP280 repeat	6.9e-18	72.9	1	306-402
726	zf-B_box	B-box zinc finger	5.6e-05	30.0	1	98-139
727	RhoGAP	RhoGAP domain	2.3e-50	180.8	1	775-947
727	DAG_PE-bind	Phorbol esters/diacylglycerol binding dom	0.0004	21.8	1	703-747
728	CN_hydrolase	Carbon-nitrogen hydrolase	0.0048	-84.5	1	25-261
729	tsp_1	Thrombospondin type 1 domain	6.9e-32	119.4	11	570-623:980-1034:1037-1089:1092-1146:1165-1220:1221-1276:1313-1364:1367-1420:1426-1479:1482-1535:1543-1593
729	Reprolysin	Reprolysin (M12B) family zinc metallo	1.3e-16	68.6	1	274-480
729	Pep_M12B_propep	Reprolysin family propeptide	4.8e-10	46.8	1	93-223
731	ig	Immunoglobulin domain	5.1e-12	53.4	3	6-99:146-235:282-373
732	ig	Immunoglobulin domain	1.6e-16	68.3	4	42-129:179-272:319-408:455-546
735	RhoGEF	RhoGEF domain	3e-10	47.5	1	165-340

471

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
737	rrm	RNA recognition motif.	1.1e-26	102.1	3	78-142:151-222:240-311
742	cadherin	Cadherin domain	3.6e-100	346.2	6	147-243:257-349:369-460:474-563:577-666:685-773
743	PGM_PMM_II	Phosphoglucomutase/phosphomannomutase, alp	0.08	-11.7	1	67-179
745	zf-C2H2	Zinc finger, C2H2 type	2.5e-108	373.3	15	130-152:158-180:186-208:214-236:242-264:270-292:298-320:326-348:354-376:382-404:410-432:438-460:488-510:516-538:544-566
746	zf-C2H2	Zinc finger, C2H2 type	9.2e-91	314.9	12	205-227:233-255:261-283:289-311:317-339:345-367:373-395:401-423:429-451:457-479:485-507:513-535
746	KRAB	KRAB box	2.3e-23	91.1	1	35-75
747	EMP24_GP25L	emp24/gp25L/p24 family	1.2e-79	278.0	1	5-201
748	acid_phosphat	Histidine acid phosphatase	2.5e-158	539.4	1	31-371
749	ArfGap	Putative GTP-ase activating protein for Arf	8.7e-60	212.0	1	527-649
749	PH	PH domain	8e-17	69.3	1	393-487
749	ank	Ankyrin repeat	4.6e-15	63.5	3	826-858:859-891:892-925
751	zf-C2H2	Zinc finger, C2H2 type	3.3e-43	157.0	6	603-625:631-653:693-715:721-743:751-773:779-801
751	KRAB	KRAB box	9.5e-20	79.0	1	342-382
753	LRR	Leucine Rich Repeat	2e-30	114.5	8	61-82:83-106:107-131:132-155:156-179:180-203:204-227:228-251
753	LRRCT	Leucine rich repeat C-terminal domain	3.6e-07	37.2	1	261-311
754	A2M	Alpha-2-macroglobulin family	3.4e-195	661.8	1	721-1469
754	A2M_N	Alpha-2-macroglobulin family N-terminal regi	1.6e-88	307.5	1	1-623
755	fibrinogen_C	Fibrinogen beta and gamma chains, C-term	4.6e-24	93.4	1	242-422
756	fn3	Fibronectin type III domain	2.1e-53	190.9	4	598-687:700-790:802-891:903-986
756	ig	Immunoglobulin domain	1.6e-49	177.9	6	43-102:137-198:242-299:332-388:424-481:514-579
758	LRR	Leucine Rich Repeat	1.2e-28	108.6	7	52-75:76-99:100-123:124-147:148-171:172-195:196-216
758	ig	Immunoglobulin domain	5.2e-07	36.7	1	301-359

472

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
758	LRRCT	Leucine rich repeat C-terminal domain	0.00013	28.8	1	240-285
759	7tm_2	7 transmembrane receptor (Secretin family)	2.3e-20	81.1	1	1009-1273
759	GPS	Latrophilin/CL-1-like GPS domain	7.1e-13	56.2	1	950-1002
759	ig	Immunoglobulin domain	3.3e-08	40.7	2	286-352:485-547
759	SEA	SEA domain	0.043	20.1	1	168-279
760	zf-C2H2	Zinc finger, C2H2 type	1.5e-85	297.6	12	107-129:135-157:163-185:191-213:219-241:247-269:275-297:303-325:331-353:359-381:387-409:415-437
764	HIT	HIT family	0.00082	-4.2	1	173-273
768	SRCR	Scavenger receptor cysteine-rich domain	2e-49	177.6	2	32-129:142-247
768	Lysyl oxidase	Lysyl oxidase	4.5e-41	149.9	1	251-359
769	Glyco_transf_8	Glycosyl transferase family 8	4.7e-06	-2.1	1	1-250
770	WD40	WD domain, G-beta repeat	4.4e-07	37.0	3	215-251:365-401:407-443
773	Cytidyltrans	Phosphatidate cytidyltransferase	8e-92	318.5	1	221-401
774	WD40	WD domain, G-beta repeat	1.5e-08	41.8	2	166-203:327-363
779	HesB-like	HesB-like domain	3.5e-36	133.6	1	49-151
780	ig	Immunoglobulin domain	0.014	22.0	2	8-57:96-155
783	vwa	von Willebrand factor type A domain	2.1e-42	154.3	1	266-440
783	Kunitz_BPTI	Kunitz/Bovine pancreatic trypsin inhibito	1.7e-18	74.8	1	540-590
783	Collagen	Collagen triple helix repeat (20 copies)	0.014	-13.0	4	2-60:61-117:118-175:181-239
784	Sterol desat	Sterol desaturase	6.4e-46	166.0	1	57-263
785	ig	Immunoglobulin domain	2e-32	121.1	4	116-176:331-391:1355-1415:1552-1613
786	adenylatekinase	Adenylate kinase	2.6e-08	-30.8	1	35-189
788	SH3	SH3 domain	6.7e-13	56.3	1	1-56
789	SH3	SH3 domain	1.6e-14	61.6	1	73-129
790	TIMP	Tissue inhibitor of metalloproteinase	1.1e-40	148.5	1	15-124
791	lectin_c	Lectin C-type domain	5.1e-06	33.3	1	162-267
792	UDPGT	UDP-glucuronosyl and UDP-glucosyl transferas	5e-237	800.8	1	1-447
794	Ubie_methyltransferan	ubiE/COQ5 methyltransferase family	6.3e-05	-96.3	1	37-241
794	PCMT	Protein-L-isoaspartate(D-aspartate) O	0.038	-104.6	1	23-192
795	7tm_1	7 transmembrane receptor (rhodopsin family)	6.9e-31	116.0	1	444-720
799	PH	PH domain	2.8e-18	74.1	1	14-112
804	ig	Immunoglobulin domain	0.0006	26.5	2	35-111:146-197
809	ig	Immunoglobulin domain	0.0014	25.4	1	109-171

473

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
811	MHC_I	Class I Histocompatibility antigen, domains	1.1e-06	4.5	1	29-205
812	ig	Immunoglobulin domain	5.4e-41	149.6	5	78-137:176-237:274-335:369-430:465-529
813	ig	Immunoglobulin domain	2.2e-103	356.8	12	295-358:393-452:1468-1530:1565-1627:1662-1724:1761-1823:1858-1926:1961-2020:2059-2120:2157-2218:2252-2313:2348-2412
814	ig	Immunoglobulin domain	2.2e-103	356.8	12	490-553:588-647:1663-1725:1760-1822:1857-1919:1956-2018:2053-2121:2156-2215:2254-2315:2352-2413:2447-2508:2543-2607
814	LRR	Leucine Rich Repeat	1.1e-25	98.8	6	58-81:82-105:106-129:130-153:154-177:186-209
814	LRRCT	Leucine rich repeat C-terminal domain	7.1e-09	42.9	1	219-280
814	LRRNT	Leucine rich repeat N-terminal domain	0.00025	27.8	1	28-56
816	Apolipoprotein	Apolipoprotein A1/A4/E family	1.6e-06	34.6	1	4-251
817	Apolipoprotein	Apolipoprotein A1/A4/E family	1.6e-06	34.6	1	4-251
819	phoslip	Phospholipase A2	3.3e-48	173.6	1	21-145
821	MR_MLE	Mandelate racemase / muconate lactonizing en	4.6e-05	-4.2	1	149-386
821	MR_MLE_N	Mandelate racemase / muconate lactonizing en	0.0031	-0.4	1	1-112
822	NAP	Nucleosome assembly protein (NAP)	1.7e-190	646.3	1	12-285
823	PP2C	Protein phosphatase 2C	6.2e-72	252.4	1	107-383
824	vwc	von Willebrand factor type C domain	3.8e-13	57.1	2	103-157:160-214
825	7tm_1	7 transmembrane receptor (rhodopsin family)	0.00045	-23.4	1	1-173
826	7tm_1	7 transmembrane receptor (rhodopsin family)	2.2e-40	147.6	1	40-287
828	RhoGAP	RhoGAP domain	1.9e-26	101.3	1	101-250
829	CUB	CUB domain	1.1e-27	105.4	1	2-102
830	CUB	CUB domain	1.1e-27	105.4	1	2-102
831	myosin_head	Myosin head (motor domain)	9.7e-15	-285.0	1	37-318
832	myosin_head	Myosin head (motor domain)	4.9e-23	-122.5	1	37-408

474

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
834	thyroglobulin_1	Thyroglobulin type-1 repeat	1.1e-20	82.1	1	316-379
834	kazal	Kazal-type serine protease inhibitor	1.5e-06	35.2	1	139-183
838	LRR	Leucine Rich Repeat	9.7e-45	162.1	12	61-84:85-108:109-132:133-156:157-180:181-204:205-228:229-252:253-276:277-300:301-324:326-349
838	LRRCT	Leucine rich repeat C-terminal domain	7.5e-09	42.8	1	359-405
838	LRRNT	Leucine rich repeat N-terminal domain	0.031	20.9	1	31-59
841	ank	Ankyrin repeat	8e-33	122.5	4	1-27:29-61:130-162:164-196
841	SAM	SAM domain (Sterile alpha motif)	0.0031	24.2	1	577-640
844	ig	Immunoglobulin domain	6.3e-39	142.8	4	53-110:150-216:255-310:350-417
845	ig	Immunoglobulin domain	5e-56	199.5	6	53-110:150-216:255-310:350-417:456-516:553-617
845	MAM	MAM domain	1.3e-52	188.2	1	753-918
847	PLA2_B	Lysophospholipase catalytic domain	4.6e-50	179.8	1	1108-1551
847	C2	C2 domain	1.6e-06	35.1	1	797-880
848	PLA2_B	Lysophospholipase catalytic domain	8.3e-53	188.9	1	357-800
848	C2	C2 domain	1.6e-06	35.1	1	46-129
851	ig	Immunoglobulin domain	3.6e-31	117.0	3	48-105:169-227:265-344
852	ig	Immunoglobulin domain	3.6e-31	117.0	3	44-101:165-223:261-340
853	ig	Immunoglobulin domain	2.8e-07	37.6	1	44-101
854	C2	C2 domain	1.3e-70	248.0	2	158-245:289-377
855	tsp_1	Thrombospondin type 1 domain	1.7e-26	101.5	6	546-596:827-881:945-995:1314-1364:1426-1471:1474-1530
855	Reprolysin	Reprolysin (M12B) family zinc metallo	1.3e-15	65.3	1	246-456
855	Pep_M12B_propep	Reprolysin family propeptide	9.2e-05	8.5	1	105-222
857	abhydrolase_2	Phospholipase/Carboxylesterase	0.051	-67.3	1	120-326
858	abhydrolase_2	Phospholipase/Carboxylesterase	0.051	-67.3	1	113-319
859	SRCR	Scavenger receptor cysteine-rich domain	3e-20	80.7	1	336-433
859	Collagen	Collagen triple helix repeat (20 copies)	2.1e-12	54.7	1	255-314
860	SRCR	Scavenger receptor cysteine-rich domain	2e-33	124.5	1	396-493
860	Collagen	Collagen triple helix repeat	9.1e-13	55.8	1	315-374

475

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		(20 copies)				
862	zf-C2H2	Zinc finger, C2H2 type	1.6e-89	310.8	12	192-214:220-242:248-270:276-298:304-326:332-354:360-382:388-410:416-438:444-466:472-494:500-523
864	zf-CCCH	Zinc finger C-x8-C-x5-C-x3-H type	1e-06	35.8	1	52-78
865	WD40	WD domain, G-beta repeat	5.7e-12	53.2	3	203-238:271-307:360-393
867	aminotran_3	Aminotransferase class-III	3.3e-98	339.7	1	76-509
868	aminotran_3	Aminotransferase class-III	6.8e-48	172.5	1	2-406
869	trypsin	Trypsin	7e-63	222.3	1	63-289
870	Glycos_transf_1	Glycosyl transferases group 1	1.8e-06	33.8	1	86-239
873	EGF	EGF-like domain	1.2e-120	414.3	16	7-43:50-81:88-119:126-157:168-199:203-234:243-279:280-311:319-350:358-389:396-427:492-523:530-561:568-599:606-637:1046-1077
873	fn3	Fibronectin type III domain	4.1e-34	126.7	3	641-722:740-823:839-921
873	sushi	Sushi domain (SCR repeat)	3.8e-05	30.5	1	433-486
875	AdoHcyase	S-adenosyl-L-homocysteine hydrolase	1.5e-280	945.4	1	81-507
878	fibrinogen_C	Fibrinogen beta and gamma chains, C-term	7.4e-54	192.3	1	146-382
879	fibrinogen_C	Fibrinogen beta and gamma chains, C-term	7.4e-54	192.3	1	146-382
880	fibrinogen_C	Fibrinogen beta and gamma chains, C-term	7.4e-54	192.3	1	146-382
883	aa_permeases	Amino acid permease	3.9e-07	-148.3	1	40-475
883	Aa_trans	Transmembrane amino acid transporter pro	0.0067	-123.4	1	42-460
884	pkinase	Protein kinase domain	9.3e-06	-52.2	1	100-659
885	lectin_c	Lectin C-type domain	0.0011	6.9	1	47-128
888	Peptidase_M20	Peptidase family M20/M25/M40	0.00043	16.2	1	55-357
889	sugar_tr	Sugar (and other) transporter	0.017	-118.8	1	1-335
891	ig	Immunoglobulin domain	7.5e-05	29.6	1	55-127
892	bromodomain	Bromodomain	6.9e-87	302.1	2	63-152:356-445
893	OLF	Olfactomedin-like domain	1.2e-120	414.2	1	220-470
894	ig	Immunoglobulin domain	7.1e-16	66.2	2	262-322:354-414
894	kazal	Kazal-type serine protease inhibitor domain	1e-09	45.7	1	88-132
894	efhand	EF hand	0.0013	25.4	1	178-206
895	aminotran_1_2	Aminotransferase class I and II	8.5e-11	49.3	1	81-416
896	LIM	LIM domain	5.4e-42	152.9	4	24-80:83-140:153-

476

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						209:212-271
896	VHP	Villin headpiece domain	7.1e-20	79.5	1	538-573
897	pkinase	Protein kinase domain	8.1e-102	351.7	1	356-613
898	pkinase	Protein kinase domain	8.1e-102	351.7	1	543-800
898	DCX	Doublecortin	5.7e-10	46.6	1	130-194
899	GST_C	Glutathione S-transferase, C-terminal domain	0.088	11.8	1	254-370
900	C1q	C1q domain	7.6e-72	252.1	1	116-241
900	Collagen	Collagen triple helix repeat (20 copies)	8.4e-06	32.7	1	37-97
902	BRCT	BRCA1 C Terminus (BRCT) domain	4e-92	319.5	6	10-93:96-183:479-570:579-666:737-823:846-944
903	BRCT	BRCA1 C Terminus (BRCT) domain	2.7e-06	34.4	1	10-93
905	LRRCT	Leucine rich repeat C-terminal domain	7.5e-09	42.8	1	37-83
905	LRR	Leucine Rich Repeat	0.0066	23.1	1	4-27
906	ig	Immunoglobulin domain	0.002	24.8	1	25-79
907	TB2_DP1_HV A22	TB2/DP1, HVA22 family	1e-34	128.7	1	2-96
908	An_peroxidase	Animal haem peroxidase	3e-193	655.4	1	770-1309
908	ig	Immunoglobulin domain	1e-34	128.8	4	224-283:320-376:409-472:533-590
908	LRR	Leucine Rich Repeat	4.7e-22	86.7	4	51-74:75-98:99-122:123-146
908	LRRCT	Leucine rich repeat C-terminal domain	8.4e-11	49.3	1	156-208
908	vwc	von Willebrand factor type C domain	7e-08	39.6	1	1439-1494
908	TILa	TILa domain	0.023	12.0	1	1438-1491
909	An_peroxidase	Animal haem peroxidase	3e-193	655.4	1	801-1340
909	ig	Immunoglobulin domain	1e-34	128.8	4	255-314:351-407:440-503:564-621
909	LRR	Leucine Rich Repeat	4.7e-22	86.7	4	82-105:106-129:130-153:154-177
909	LRRCT	Leucine rich repeat C-terminal domain	8.4e-11	49.3	1	187-239
909	vwc	von Willebrand factor type C domain	7e-08	39.6	1	1470-1525
909	TILa	TILa domain	0.023	12.0	1	1469-1522
910	An_peroxidase	Animal haem peroxidase	3e-193	655.4	1	663-1202
910	ig	Immunoglobulin domain	3.2e-24	93.9	3	201-260:297-353:386-449
910	LRR	Leucine Rich Repeat	2.6e-18	74.3	4	51-74:75-98:99-122:123-146
910	vwc	von Willebrand factor type C domain	7e-08	39.6	1	1332-1387
910	TILa	TILa domain	0.023	12.0	1	1331-1384
911	EGF	EGF-like domain	3.1e-50	180.3	9	47-99:106-141:172-203:210-245:574-605:823-854:861-892:901-933:940-971

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
911	laminin_G	Laminin G domain	0.0002	25.1	2	275-401:663-788
914	cNMP_binding	Cyclic nucleotide-binding domain	1.5e-65	231.2	2	152-240:270-364
914	RIIa	Regulatory subunit of type II PKA R-subu	4.8e-13	56.8	1	25-62
915	DIL	DIL domain	6.6e-40	146.0	1	214-323
915	PDZ	PDZ domain (Also known as DHR or GLGF)	2e-12	54.7	1	555-639
916	lipoygenase	Lipoygenase	3.3e-193	655.3	1	121-648
916	PLAT	PLAT/LH2 domain	1.6e-29	111.5	1	2-111
917	PLAT	PLAT/LH2 domain	1.6e-29	111.5	1	2-111
917	lipoygenase	Lipoygenase	0.00053	-342.4	1	91-294
918	PLAT	PLAT/LH2 domain	1.6e-29	111.5	1	2-111
918	lipoygenase	Lipoygenase	4e-06	-297.4	1	121-323
926	Aa_trans	Transmembrane amino acid transporter protein	1.3e-138	473.9	1	114-517
927	EGF	EGF-like domain	5.8e-36	132.9	6	29-57:60-88:95-128:135-171:178-209:216-247
930	DUF6	Integral membrane protein DUF6	0.00017	28.3	2	8-129:147-277
933	Peptidase_M24	metallopeptidase family M24	2.1e-69	244.0	1	87-326
938	PDZ	PDZ domain (Also known as DHR or GLGF)	1.8e-20	81.4	1	93-174
938	L27	L27 domain	6.5e-16	66.3	1	13-68
940	rrm	RNA recognition motif.	2.7e-46	167.2	4	61-128:186-253:339-406:456-524
941	EGF	EGF-like domain	1.9e-18	74.7	5	66-92:99-133:139-173:179-218:224-259
941	MAM	MAM domain	1.7e-13	58.3	1	403-547
942	EGF	EGF-like domain	1.9e-18	74.7	5	71-97:104-138:144-178:184-223:229-264
942	MAM	MAM domain	1.7e-13	58.3	1	408-552
943	PHD	PHD-finger	2.9e-10	47.5	1	85-128
943	bromodomain	Bromodomain	8.2e-10	46.0	1	149-235
943	zf-MYND	MYND finger	7e-07	36.3	1	977-1011
943	PWWP	PWWP domain	7.5e-06	32.9	1	269-340
944	PHD	PHD-finger	2.9e-10	47.5	1	85-128
944	bromodomain	Bromodomain	8.2e-10	46.0	1	149-235
944	PWWP	PWWP domain	7.5e-06	32.9	1	269-340
945	PHD	PHD-finger	2.9e-10	47.5	1	85-128
945	bromodomain	Bromodomain	8.2e-10	46.0	1	149-235
945	zf-MYND	MYND finger	7e-07	36.3	1	1023-1057
945	PWWP	PWWP domain	7.5e-06	32.9	1	269-340
946	PHD	PHD-finger	2.9e-10	47.5	1	90-133
946	bromodomain	Bromodomain	8.2e-10	46.0	1	154-240
946	zf-MYND	MYND finger	7e-07	36.3	1	1028-1062
946	PWWP	PWWP domain	7.5e-06	32.9	1	274-345
950	ion_trans	Ion transport protein	3.5e-19	77.1	1	345-518
951	Reprolysin	Reprolysin (M12B) family zinc metallo	3e-88	306.6	1	210-409
951	Pep_M12B_propep	Reprolysin family propeptide	1.3e-31	118.4	1	80-198

478

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
951	disintegrin	Disintegrin	2.5e-23	90.9	1	426-501
953	ank	Ankyrin repeat	2e-46	167.6	7	151-183:184-215:216-248:250-282:283-328:329-361:362-401
954	interferon	Interferon alpha/beta domain	1.8e-17	71.5	1	16-171
956	adh_short	short chain dehydrogenase	1.3e-07	21.8	1	31-188
958	acid_phosphat	Histidine acid phosphatase	1.7e-58	207.7	1	30-381
959	serpin	Serpin (serine protease inhibitor)	6.4e-179	607.8	1	1-329
960	serpin	Serpin (serine protease inhibitor)	9.1e-200	677.1	1	47-397
961	serpin	Serpin (serine protease inhibitor)	3.2e-200	678.5	1	47-397
962	serpin	Serpin (serine protease inhibitor)	1.2e-203	689.9	1	47-397
964	Reprolysin	Reprolysin (M12B) family zinc metallo	5.8e-96	332.2	1	232-426
964	Pep_M12B_pr opep	Reprolysin family propeptide	4.4e-41	149.9	1	112-220
964	disintegrin	Disintegrin	2.5e-09	44.5	1	444-517
965	Uteroglobin	Uteroglobin family	1.4e-05	31.8	1	1-88
966	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) family	5.7e-92	319.0	1	48-483
967	C1q	C1q domain	6.1e-44	159.4	1	73-202
970	ig	Immunoglobulin domain	1.6e-06	35.1	2	41-124:156-230
970	zf-CCHC	Zinc knuckle	5.7e-05	30.0	1	523-540
971	pentaxin	Pentaxin family	8.1e-22	85.9	1	281-479
973	bZIP	bZIP transcription factor	0.024	19.0	1	622-686
974	WD40	WD domain, G-beta repeat	0.003	24.3	4	37-72:77-113:122-156:211-247
975	ion_trans	Ion transport protein	0.0031	24.2	1	248-408
976	ion_trans	Ion transport protein	0.0031	24.2	1	322-482
977	zf-C2H2	Zinc finger, C2H2 type	2.5e-55	197.2	35	4-27:108-131:162-185:243-266:439-462:470-492:600-623:843-866:886-908:925-948:1030-1053:1114-1137:1193-1216:1265-1288:1312-1335:1369-1392:1470-1493:1515-1538:1577-1600:1660-1683:1697-1720:1767-1790:1846-1869:1892-1914:1968-1990:2051-2073:2085-2107:2114-

479

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						2137:2143- 2166:2251- 2274:2280- 2303:2314- 2336:2360- 2382:2388- 2411:2474-2496
980	trypsin	Trypsin	7.9e-18	72.7	1	155-326
980	PDZ	PDZ domain (Also known as DHR or GLGF)	8e-12	52.7	1	332-427
980	kazal	Kazal-type serine protease inhibitor domain	3.7e-05	30.6	1	63-117
981	asp	Eukaryotic aspartyl protease	8.1e-104	358.3	1	19-421
984	Zn_carbOpept	Zinc carboxypeptidase	2e-114	393.5	1	50-332
985	Zn_carbOpept	Zinc carboxypeptidase	2e-114	393.5	1	50-332
986	NifU_N	NifU-like N terminal domain	4.2e-80	279.5	1	34-160
988	UPAR_LY6	u-PAR/Ly-6 domain	1.8e-05	31.6	1	28-110
990	zf-C2H2	Zinc finger, C2H2 type	1.4e-12	55.2	3	53-78:87-114:120-144
991	pkinase	Protein kinase domain	8.6e-90	311.7	1	20-312
992	spectrin	Spectrin repeat	6.6e-26	99.5	7	17-121:124-226:229-340:343-449:452-556:781-888:891-999
994	C1q	C1q domain	2.1e-31	117.8	1	160-284
994	Collagen	Collagen triple helix repeat (20 copies)	0.00022	22.3	1	76-135
995	Allantoicase	Allantoicase repeat	8.7e-122	418.0	2	1-136:159-319
996	ig	Immunoglobulin domain	4.9e-11	50.1	3	37-151:182-243:275-335
997	RasGEF	RasGEF domain	1.7e-88	307.4	1	999-1184
997	RhoGEF	RhoGEF domain	8.2e-68	238.7	1	247-428
997	PH	PH domain	2.3e-35	130.9	2	23-133:460-588
997	RasGEFN	Guanine nucleotide exchange factor for Ras-1	4.9e-18	73.3	1	633-688
997	IQ	IQ calmodulin-binding motif	0.012	22.2	1	206-226
999	K_tetra	K ⁺ channel tetramerisation domain	6e-31	116.2	1	24-126
1002	PHD	PHD-finger	1.9e-17	71.4	1	185-233
1002	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.00078	26.2	1	108-156
1003	WD40	WD domain, G-beta repeat	1.8e-24	94.7	6	768-802:959-992:1070-1104:1110-1145:1151-1185:1191-1225
1004	ZZ	Zinc finger, ZZ type	4.6e-11	50.2	1	3-48
1004	zf-C2H2	Zinc finger, C2H2 type	0.012	22.2	1	78-101
1006	C2	C2 domain	9.6e-05	29.2	1	304-394
1007	IBN_NT	Importin-beta N-terminal domain	9.5e-28	105.6	1	22-101
1009	ArfGap	Putative GTP-ase activating protein for Arf	1.4e-35	131.6	1	250-373

480

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1009	PH	PH domain	1.7e-14	61.6	1	136-227
1009	ank	Ankyrin repeat	2e-11	51.4	2	411-446:447-479
1009	SH3	SH3 domain	1.7e-10	48.3	1	881-938
1011	ig	Immunoglobulin domain	1.2e-48	175.1	6	80-148:183-242:281-342:379-440:474-535:570-634
1015	efhand	EF hand	3.7e-26	100.3	4	29-57:65-93:102-130:138-166
1018	7tm_1	7 transmembrane receptor (rhodopsin family)	3.7e-76	266.4	1	87-350
1019	LRR	Leucine Rich Repeat	2.9e-41	150.5	14	82-105:106-129:133-157:158-181:182-205:206-229:251-272:329-352:377-399:403-426:427-444:463-486:537-558:559-582
1021	RasGEF	RasGEF domain	1e-47	172.0	1	907-1092
1021	PDZ	PDZ domain (Also known as DHR or GLGF)	4.2e-17	70.2	1	580-661
1021	cNMP_binding	Cyclic nucleotide-binding domain	3.8e-13	57.1	1	345-435
1021	RA	Ras association (RalGDS/AF-6) domain	1.3e-05	32.1	1	799-885
1022	RasGEF	RasGEF domain	1e-47	172.0	1	857-1042
1022	PDZ	PDZ domain (Also known as DHR or GLGF)	4.2e-17	70.2	1	530-611
1022	cNMP_binding	Cyclic nucleotide-binding domain	3.8e-13	57.1	1	295-385
1022	RA	Ras association (RalGDS/AF-6) domain	1.3e-05	32.1	1	749-835
1026	Ricin_B_lectin	QXW lectin repeat	1.3e-11	52.1	3	134-172:187-225:226-265
1027	SCF	Stem cell factor	2.4e-119	409.9	1	1-216
1028	cadherin	Cadherin domain	1.9e-75	264.0	4	50-141:155-250:264-366:379-470
1029	cadherin	Cadherin domain	1.4e-78	274.5	4	50-141:155-250:264-366:379-470
1030	PH	PH domain	1.2e-10	48.8	1	522-624
1031	Renal_dipeptase	Renal dipeptidase	1.3e-73	258.0	1	54-377
1032	aa_permeases	Amino acid permease	3.9e-07	-148.3	1	40-475
1032	Aa_trans	Transmembrane amino acid transporter pro	0.0067	-123.4	1	42-460
1033	FTHFS	Formate--tetrahydrofolate ligase	0	1367.2	1	360-979
1033	THF_DHG_C YH	Tetrahydrofolate dehydrogenase/cyclohyd	1.5e-07	21.3	1	68-180
1033	THF_DHG_C YH_C	Tetrahydrofolate dehydrogenase/cyclohyd	3.7e-05	-45.5	1	182-329
1035	RhoGEF	RhoGEF domain	9.1e-26	99.0	1	778-962
1035	PDZ	PDZ domain (Also known as DHR or GLGF)	4.2e-12	53.6	1	47-122
1035	PH	PH domain	0.081	19.5	1	1006-1119

481

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1037	PH	PH domain	2.4e-10	47.8	1	17-124
1037	efhand	EF hand	2.7e-08	41.0	2	138-166:174-202
1039	7tm_1	7 transmembrane receptor (rhodopsin family)	3.9e-22	87.0	1	40-289
1040	tsp_3	Thrombospondin type 3 repeat	1.1e-22	88.9	9	404-418:440-454:463-477:499-513:522-536:537-551:560-574:600-614:615-627
1040	TSPN	Thrombospondin N-terminal -like domain	2.3e-05	22.9	1	1-101
1042	PTR2	POT family	7.4e-85	295.3	1	103-471
1043	FH2	Formin Homology 2 Domain	4e-105	362.7	1	595-1038
1044	zf-C2H2	Zinc finger, C2H2 type	4.3e-140	478.8	19	114-136:142-164:170-192:198-220:226-248:254-276:282-304:310-332:338-360:366-388:394-416:422-444:450-472:478-500:506-528:534-556:562-584:590-612:618-640
1044	KRAB	KRAB box	6.4e-27	102.9	1	8-48
1044	zf-BED	BED zinc finger	0.099	10.5	2	431-473:603-641
1046	PA	PA domain	4.1e-20	80.3	1	155-255
1048	TIG	IPT/TIG domain	5.9e-57	202.6	3	803-893:895-980:983-1092
1048	PSI	Plexin repeat	7.4e-26	99.3	2	468-519:759-801
1048	Sema	Sema domain	1.6e-11	-3.7	1	34-449
1049	BTB	BTB/POZ domain	1.7e-26	101.4	1	20-124
1050	ABC_tran	ABC transporter	9.9e-37	135.5	1	26-217
1051	ZZ	Zinc finger, ZZ type	4.6e-11	50.2	1	3-48
1051	zf-C2H2	Zinc finger, C2H2 type	0.012	22.2	1	78-101
1052	ig	Immunoglobulin domain	1.2e-11	52.2	2	34-110:150-204
1053	CUB	CUB domain	2.5e-12	54.4	1	156-260
1053	WSC	WSC domain	0.002	18.6	1	71-142
1054	ig	Immunoglobulin domain	0.0026	24.4	1	36-113
1055	MHC_I	Class I Histocompatibility antigen, domains	2.4e-146	499.6	1	25-203
1055	ig	Immunoglobulin domain	8.5e-08	39.3	1	220-285
1057	LBP_BPI_CE TP_C	LBP / BPI / CETP family, C-terminal do	0.00076	-0.8	1	217-444
1062	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	1.8e-44	161.2	1	4-181
1064	PDZ	PDZ domain (Also known as DHR or GLGF)	4.8e-71	249.5	5	1-84:209-297:310-393:409-490:694-775
1065	PID	Phosphotyrosine interaction domain (PTB/PID)	1.1e-44	161.8	1	42-168
1067	pkinase	Protein kinase domain	2.8e-73	256.8	1	12-272
1068	lipocalin	Lipocalin / cytosolic fatty-acid binding pr	5.6e-37	136.3	1	38-185
1069	lactamase_B	Metallo-beta-lactamase	3e-35	130.6	1	7-172

482

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
		superfamily				
1070	annexin	Annexin	2e-128	440.1	4	57-124:128-196:212-280:288-355
1071	SNF	Sodium:neurotransmitter symporter family	0	1202.5	1	44-574
1072	ig	Immunoglobulin domain	0.0008	26.1	1	38-122
1073	Glypican	Glypican	2.1e-291	981.5	1	3-566
1074	PAP_assoc	PAP/25A associated domain	4.2e-12	53.7	1	490-549
1074	rrm	RNA recognition motif.	7.2e-08	39.6	1	58-123
1075	Glyco_transf_29	Glycosyltransferase family 29	3.6e-69	243.2	1	213-507
1078	A2M	Alpha-2-macroglobulin family	3.4e-195	661.8	1	721-1469
1078	A2M_N	Alpha-2-macroglobulin family N-terminal regi	1.6e-88	307.5	1	1-623
1079	A2M_N	Alpha-2-macroglobulin family N-terminal regi	4.7e-90	312.6	1	14-636
1080	A2M_N	Alpha-2-macroglobulin family N-terminal regi	1.5e-38	141.5	1	1-563
1081	A2M	Alpha-2-macroglobulin family	1.3e-200	679.9	1	721-1469
1081	A2M_N	Alpha-2-macroglobulin family N-terminal regi	1.6e-88	307.5	1	1-623
1082	A2M_N	Alpha-2-macroglobulin family N-terminal regi	4.7e-90	312.6	1	1-623
1083	COesterase	Carboxylesterase	2.1e-155	529.7	1	6-547
1084	EGF	EGF-like domain	9.5e-90	311.6	18	192-219:404-431:631-666:878-914:920-956:962-997:1003-1037:1043-1078:1084-1119:1125-1160:1166-1201:1207-1243:1249-1285:1291-1328:1429-1466:1472-1507:1626-1661:1667-1706
1084	TB	TB domain	1.8e-78	274.1	4	567-610:688-729:1358-1401:1535-1577
1086	fn3	Fibronectin type III domain	5.9e-95	328.9	5	373-459:501-587:602-685:700-786:802-888
1086	ig	Immunoglobulin domain	3e-24	94.0	4	168-232:285-347:1133-1191:1349-1409
1087	zf-C2H2	Zinc finger, C2H2 type	4.6e-33	123.3	4	161-183:189-211:217-239:245-267
1087	KRAB	KRAB box	1.9e-24	94.6	1	14-54
1088	KRAB	KRAB box	1.9e-24	94.6	1	14-54
1088	zf-C2H2	Zinc finger, C2H2 type	1e-07	39.1	1	161-183

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1089	Keratin_B2	Keratin, high sulfur B2 protein	2.5e-17	71.0	1	2-153
1090	Keratin_B2	Keratin, high sulfur B2 protein	0.0059	-33.4	1	2-76
1091	Keratin_B2	Keratin, high sulfur B2 protein	1.9e-06	21.1	2	2-108:111-205
1092	abhydrolase	alpha/beta hydrolase fold	1.6e-12	55.0	1	111-390
1093	abhydrolase	alpha/beta hydrolase fold	1.6e-12	55.0	1	171-450
1094	7tm_3	7 transmembrane receptor	5.5e-08	10.8	1	22-271
1096	lectin_c	Lectin C-type domain	2.3e-25	97.7	1	100-208
1097	lectin_c	Lectin C-type domain	6.5e-27	102.8	1	100-208
1098	7tm_1	7 transmembrane receptor (rhodopsin family)	1.7e-41	151.3	1	41-290
1099	SEA	SEA domain	0.00037	27.2	1	330-447
1100	ig	Immunoglobulin domain	5e-11	50.1	3	146-203:245-295:331-405
1101	An_peroxidase	Animal haem peroxidase	2.7e-194	658.9	1	726-1265
1101	ig	Immunoglobulin domain	4.4e-36	133.3	4	248-307:344-400:433-490:525-582
1101	LRR	Leucine Rich Repeat	3.1e-25	97.3	5	51-74:75-98:99-122:123-146:147-170
1101	LRRCT	Leucine rich repeat C-terminal domain	8.4e-11	49.3	1	180-232
1101	vwc	von Willebrand factor type C domain	7e-08	39.6	1	1395-1450
1101	TILa	TILa domain	0.023	12.0	1	1394-1447
1102	An_peroxidase	Animal haem peroxidase	2.7e-194	658.9	1	702-1241
1102	ig	Immunoglobulin domain	4.4e-36	133.3	4	224-283:320-376:409-466:501-558
1102	LRR	Leucine Rich Repeat	3.2e-21	83.9	4	51-74:75-98:99-122:123-146
1102	LRRCT	Leucine rich repeat C-terminal domain	8.4e-11	49.3	1	156-208
1102	vwc	von Willebrand factor type C domain	7e-08	39.6	1	1371-1426
1102	TILa	TILa domain	0.023	12.0	1	1370-1423
1113	pkinase	Protein kinase domain	3e-45	163.8	1	194-468
1117	ig	Immunoglobulin domain	5.8e-17	69.8	4	30-87:127-186:281-337:375-434
1118	ig	Immunoglobulin domain	0.00012	28.9	2	42-98:136-195
1119	IBN_NT	Importin-beta N-terminal domain	3.4e-23	90.5	1	28-100
1120	ank	Ankyrin repeat	7.7e-21	82.7	2	920-952:953-985
1120	SH3	SH3 domain	6.1e-15	63.1	1	1022-1079
1122	TPR	TPR Domain	6.4e-09	43.1	3	124-157:158-191:192-225
1124	ank	Ankyrin repeat	2.9e-46	167.1	6	31-63:64-96:97-129:130-162:163-195:196-228
1125	ank	Ankyrin repeat	3.4e-38	140.3	5	31-63:64-96:97-129:130-162:163-195
1129	F5_F8_type_C	F5/8 type C domain	1.4e-54	194.8	1	34-174
1129	laminin_G	Laminin G domain	1.4e-07	38.6	1	212-344

484

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1130	F5_F8_type_C	F5/8 type C domain	1.4e-54	194.8	1	34-174
1130	laminin_G	Laminin G domain	3.1e-44	160.4	4	212-344:398-525:821-943:1046-1179
1130	EGF	EGF-like domain	9.1e-07	35.9	2	551-583:962-996
1131	Glycos_transf_2	Glycosyl transferase	6.5e-31	116.1	1	155-341
1131	Ricin_B_lectin	QXW lectin repeat	0.00059	26.6	2	467-507:558-596
1133	pkinase	Protein kinase domain	1.7e-48	174.5	1	11-347
1135	C2	C2 domain	1.1e-42	155.2	2	7-88:135-216
1135	RasGAP	GTPase-activator protein for Ras-like GTPase	5.2e-34	126.4	1	323-513
1135	PH	PH domain	5.8e-08	39.9	1	567-673
1135	BTK	BTK motif	9.2e-05	28.9	1	675-711
1137	MAM	MAM domain	1.1e-22	88.9	1	452-593
1137	EGF	EGF-like domain	3.5e-15	63.9	5	60-86:123-157:163-197:203-242:248-283
1143	7tm_1	7 transmembrane receptor (rhodopsin family)	0.00045	-23.4	1	1-173
1144	7tm_1	7 transmembrane receptor (rhodopsin family)	2.2e-40	147.6	1	40-287
1147	IL1	Interleukin-1 / 18	4.3e-21	83.5	1	12-152
1148	filament	Intermediate filament protein	5e-101	349.0	1	1-299
1150	MBOAT	MBOAT family	1.4e-06	-27.4	1	130-323
1151	filament	Intermediate filament protein	2.1e-116	400.1	1	131-412
1152	Peptidase_M10	Matrixin	4.4e-84	292.8	1	36-202
1152	hemopexin	Hemopexin	6e-37	136.2	4	231-273:275-317:322-369:371-411
1153	Peptidase_M10	Matrixin	4.4e-84	292.8	1	36-202
1153	hemopexin	Hemopexin	6e-37	136.2	4	231-273:275-317:322-369:371-411
1155	LBP_BPI_CE_TP_C	LBP / BPI / CETP family, C-terminal do	3.1e-30	113.9	1	242-478
1155	LBP_BPI_CE_TP	LBP / BPI / CETP family, N-terminal do	3.3e-22	87.2	1	26-240
1156	HMG_box	HMG (high mobility group) box	3.1e-31	117.2	1	85-153
1159	DNA_ligase	ATP dependent DNA ligase domain	3.7e-57	203.3	1	480-645
1159	zf-PARP	Poly(ADP-ribose) polymerase and DNA-Ligase	8.5e-52	185.5	1	93-185
1160	serpin	Serpin (serine protease inhibitor)	7.7e-150	511.2	1	3-425
1167	ig	Immunoglobulin domain	3.4e-16	67.2	3	42-96:135-197:237-297
1169	lectin_c	Lectin C-type domain	2e-18	74.6	1	131-231
1171	WD40	WD domain, G-beta repeat	4.4e-80	279.5	8	224-260:280-316:321-357:363-398:404-440:446-491:497-533:539-574
1172	MBOAT	MBOAT family	1.6e-08	6.7	1	488-777
1172	ig	Immunoglobulin domain	2.9e-08	40.9	2	42-99:139-198

485

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1173	MBOAT	MBOAT family	1.6e-08	6.7	1	488-777
1173	ig	Immunoglobulin domain	2.9e-08	40.9	2	42-99:139-198
1174	MBOAT	MBOAT family	5.1e-65	229.4	1	130-373
1175	PTE	Phosphotriesterase family	1.4e-90	314.4	1	6-233
1183	PS_Dcarboxylase	Phosphatidylserine decarboxylase	6.5e-45	162.6	1	232-467
1184	TSC22	TSC-22/dip/bun family	1.3e-40	148.4	1	124-183
1188	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	5.1e-08	-71.7	1	132-680
1188	Peptidase_S9	Prolyl oligopeptidase family	1.7e-06	35.0	1	684-764
1189	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	5.1e-08	-71.7	1	132-680
1189	Peptidase_S9	Prolyl oligopeptidase family	1.7e-06	35.0	1	684-764
1190	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	3.8e-07	-94.7	1	132-667
1190	Peptidase_S9	Prolyl oligopeptidase family	1.7e-06	35.0	1	671-751
1191	Ribosomal_S25	S25 ribosomal protein	6.5e-66	232.4	1	1-100
1193	ank	Ankyrin repeat	1.2e-239	809.5	27	49-81:82-114:115-147:148-180:181-213:214-246:247-279:280-313:314-346:347-379:380-412:431-463:464-496:497-557:558-591:593-625:626-658:660-692:696-728:729-761:762-797:798-827:830-864:865-897:898-931:932-964:968-1000
1194	trypsin	Trypsin	2.5e-18	74.3	1	166-342
1196	vwc	von Willebrand factor type C domain	0.043	12.4	2	50-105:108-163
1197	7tm_1	7 transmembrane receptor (rhodopsin family)	1.2e-28	108.6	1	46-295
1198	MethyltransfD12	D12 class N6 adenine-specific DNA met	0.0057	-49.7	1	30-153
1199	lipocalin	Lipocalin / cytosolic fatty-acid binding pr	1.3e-22	88.6	1	32-176
1200	tRNA-synt_2	tRNA synthetases class II (D, K and N)	7.4e-91	315.3	1	135-473
1200	tRNA_anti	OB-fold nucleic acid binding domain	7.3e-11	49.5	1	44-118
1202	FAD_binding_2	FAD binding domain	8.6e-09	-83.1	1	5-162
1203	RasGEF	RasGEF domain	1.9e-16	68.1	1	211-412
1204	KH-domain	KH domain	1.9e-50	181.0	3	17-63:101-150:265-313
1206	transket_pyr	Transketolase, pyridine binding domai	4e-74	259.7	1	14-191
1206	transketolase_C	Transketolase, C-terminal domain	5e-55	196.2	1	208-331
1207	Calsequestrin	Calsequestrin	1.7e-	1001.7	1	1-390

486

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
			297			
1210	ig	Immunoglobulin domain	1.1e-13	58.9	2	35-112:154-228
1213	cadherin	Cadherin domain	9e-81	281.8	6	33-126:140-235:249-343:357-448:462-558:576-667
1214	calreticulin	Calreticulin family	2.7e-206	698.7	1	21-317
1221	Osteopontin	Osteopontin	4.6e-173	588.4	1	1-279
1222	serpin	Serpin (serine protease inhibitor)	2.4e-155	529.5	1	80-443
1223	ig	Immunoglobulin domain	4.8e-15	63.4	2	31-101:252-303
1225	DNA_topoisoI V	DNA gyrase/topoisomerase IV, subunit A	3.7e-180	611.9	1	653-1120
1225	DNA_gyraseB	DNA gyrase B	1.3e-56	201.6	1	210-370
1225	HATPase_c	Histidine kinase-, DNA gyrase B-, and H	1.8e-13	58.2	1	16-164
1226	AMP-binding	AMP-binding enzyme	3.6e-80	279.7	1	105-539
1227	PCI	PCI domain	0.016	18.5	1	26-117
1228	Clq	Clq domain	5.9e-45	162.8	1	73-202
1230	ank	Ankyrin repeat	3.6e-215	728.2	28	7-39:40-72:86-147:148-180:181-213:214-246:247-279:280-312:313-346:347-379:380-412:413-445:464-496:497-529:530-590:591-621:626-658:659-691:693-725:729-761:762-794:795-827:832-862:864-897:899-931:932-965:966-998:1002-1034
1231	LBP_BPI_CE TP_C	LBP / BPI / CETP family, C-terminal do	9.4e-24	92.3	1	242-470
1231	LBP_BPI_CE TP	LBP / BPI / CETP family, N-terminal do	3.3e-22	87.2	1	26-240
1232	LBP_BPI_CE TP_C	LBP / BPI / CETP family, C-terminal do	3.1e-22	87.3	1	242-470
1232	LBP_BPI_CE TP	LBP / BPI / CETP family, N-terminal do	3.3e-22	87.2	1	26-240
1233	LBP_BPI_CE TP_C	LBP / BPI / CETP family, C-terminal do	9.4e-32	118.9	1	242-478
1233	LBP_BPI_CE TP	LBP / BPI / CETP family, N-terminal do	3.3e-22	87.2	1	26-240
1237	ig	Immunoglobulin domain	2.8e-30	114.0	3	28-86:127-184:219-277
1237	fn3	Fibronectin type III domain	2.6e-28	107.5	2	299-385:396-481
1238	Nuf2	Nuf2 family	8.7e-104	358.2	1	1-148
1240	Sema	Sema domain	2.2e-177	602.7	1	59-496
1243	rrm	RNA recognition motif.	0.05	15.7	1	17-93
1247	EGF	EGF-like domain	4.8e-56	199.6	17	105-135:148-

487

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
						178:191-221:234-264:277-307:320-350:364-396:409-439:452-482:495-525:538-568:581-611:624-656:669-699:712-742:755-785:798-828
1249	IBR	IBR domain	0.069	5.8	2	36-104:111-167
1251	Aa_trans	Transmembrane amino acid transporter protein	3.7e-46	166.8	1	52-394
1252	Aa_trans	Transmembrane amino acid transporter protein	1.3e-65	231.4	1	45-419
1254	FGF	Fibroblast growth factor	1.7e-37	138.0	1	36-166
1255	LRR	Leucine Rich Repeat	0.0019	24.9	4	49-70:71-92:94-115:116-137
1256	RPE65	Retinal pigment epithelial membrane protein	5.8e-83	289.0	1	35-579
1257	RPE65	Retinal pigment epithelial membrane protein	4.7e-82	286.0	1	24-561
1258	ig	Immunoglobulin domain	3.1e-15	64.1	2	39-97:128-189
1261	serpin	Serpin (serine protease inhibitor)	1.9e-56	200.9	1	23-423
1263	arf	ADP-ribosylation factor family	7.9e-09	-6.8	1	9-182
1264	PAP2	PAP2 superfamily	3e-11	50.8	1	95-241
1265	SRCR	Scavenger receptor cysteine-rich domain	1.3e-128	440.7	5	35-128:136-227:232-329:360-459:477-574
1266	SRCR	Scavenger receptor cysteine-rich domain	1.3e-128	440.7	5	35-128:136-227:232-329:360-459:477-574
1270	Armadillo_seg	Armadillo/beta-catenin-like repeat	1.4e-05	32.0	4	53-93:546-586:633-673:675-716
1273	pkinase	Protein kinase domain	8e-77	268.6	1	103-387
1275	Reprolysin	Reprolysin (M12B) family zinc metallo	3e-88	306.6	1	227-426
1275	Pep_M12B_propep	Reprolysin family propeptide	1.3e-31	118.4	1	97-215
1275	disintegrin	Disintegrin	2.5e-23	90.9	1	443-518
1277	ank	Ankyrin repeat	2.6e-17	70.9	2	301-339:340-373
1278	Peptidase_M1	Peptidase family M1	2.6e-112	386.5	1	98-506
1284	Aa_trans	Transmembrane amino acid transporter protein	1.4e-31	118.3	1	4-407
1285	UPF0083	Uncharacterised protein family (UPF0083)	1.9e-05	14.5	1	73-213
1288	LRR	Leucine Rich Repeat	1.3e-23	91.9	7	66-89:90-113:114-137:138-161:163-186:187-210:211-233
1288	ig	Immunoglobulin domain	2.7e-07	37.7	1	314-372
1288	LRRCT	Leucine rich repeat C-terminal domain	5.6e-05	30.0	1	252-297
1290	LRR	Leucine Rich Repeat	2.2e-12	54.6	3	61-84:85-108:110-132
1291	DAGKc	Diacylglycerol kinase catalytic domain	0.063	-14.5	1	74-220

488

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1292	ig	Immunoglobulin domain	6.7e-10	46.3	2	48-124:161-219
1293	ig	Immunoglobulin domain	6.7e-10	46.3	2	48-124:161-219
1295	C1q	C1q domain	1.4e-48	174.8	1	72-198
1296	7tm_1	7 transmembrane receptor (rhodopsin family)	2.5e-24	94.3	1	49-332
1297	Plectin	Plectin repeat	1.6e-86	300.8	6	2734-2778:2808-2852:2897-2939:3003-3042:3043-3087:3119-3163
1297	CH	Calponin homology (CH) domain	1.6e-72	254.3	2	213-316:329-433
1297	spectrin	Spectrin repeat	0.029	8.2	1	889-994
1298	Plectin	Plectin repeat	1.6e-86	300.8	6	2746-2790:2820-2864:2909-2951:3015-3054:3055-3099:3131-3175
1298	CH	Calponin homology (CH) domain	3.1e-69	243.4	2	213-328:341-445
1298	spectrin	Spectrin repeat	0.029	8.2	1	901-1006
1306	MAM	MAM domain	5.8e-49	176.1	1	422-595
1306	ig	Immunoglobulin domain	5.4e-18	73.2	3	26-93:132-191:228-287
1308	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal doma	1.6e-49	178.0	1	618-769
1308	Acyl-CoA_dh_M	Acyl-CoA dehydrogenase, middle domain	1.4e-06	15.3	1	505-614
1309	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal doma	1.6e-49	178.0	1	600-751
1309	Acyl-CoA_dh_M	Acyl-CoA dehydrogenase, middle domain	1.4e-06	15.3	1	487-596
1311	IQ	IQ calmodulin-binding motif	0.00039	27.2	2	715-735:738-758
1312	SAM	SAM domain (Sterile alpha motif)	3.9e-13	57.1	2	304-369:382-446
1314	HECT	HECT-domain (ubiquitin-transferase)	5.3e-196	664.5	1	2002-2309
1315	PAP2	PAP2 superfamily	7.8e-28	105.9	1	56-218
1316	PAP2	PAP2 superfamily	1.6e-32	121.5	1	88-236
1317	ig	Immunoglobulin domain	2.7e-07	37.6	1	41-116
1321	LRR	Leucine Rich Repeat	1.9e-66	234.2	20	145-168:169-194:195-217:240-265:266-285:287-310:311-336:337-356:358-381:382-407:408-427:429-452:453-478:479-498:500-523:524-549:550-569:571-594:595-620:621-644
1321	LRRNT	Leucine rich repeat N-terminal domain	0.0027	24.4	1	115-143
1322	ig	Immunoglobulin domain	3.6e-14	60.5	3	34-120:157-215:267-321

TABLE 4A

SEQ ID	Model	Description	E-value	Score	Repeats	Position
1323	ig	Immunoglobulin domain	7.8e-06	32.8	3	34-120:157-215:267-313
1324	tsp_1	Thrombospondin type 1 domain	0.00039	27.2	1	37-81
1328	SRCR	Scavenger receptor cysteine-rich domain	1.5e-171	583.3	5	14-111:188-285:300-397:405-503:638-730
1331	efhand	EF hand	1.5e-06	35.2	3	12-40:48-76:85-113
1333	wnt	wnt family	6.8e-205	694.1	1	40-365
1336	zf-MIZ	MIZ zinc finger	3.2e-32	120.5	1	323-375
1336	SAP	SAP domain	2.4e-05	31.2	1	11-45
1337	FA_desaturase	Fatty acid desaturase	2.1e-76	267.3	1	71-296
1338	Retrotrans_gag	Retrotransposon gag protein	0.097	8.7	1	200-300
1340	actin	Actin	1.9e-61	217.5	1	1-367
1341	ion_trans	Ion transport protein	0.01	22.5	1	117-302
1343	fn3	Fibronectin type III domain	7.3e-33	122.6	2	394-480:492-578
1343	ig	Immunoglobulin domain	1.1e-23	92.1	3	124-182:224-281:316-372
1344	ig	Immunoglobulin domain	5e-56	199.5	6	53-110:150-216:255-310:350-417:456-516:553-617
1344	MAM	MAM domain	1.3e-52	188.2	1	753-918
1345	ig	Immunoglobulin domain	5.9e-05	29.9	1	186-255
1345	kazal	Kazal-type serine protease inhibitor domain	0.00028	27.6	1	121-168
1348	ig	Immunoglobulin domain	3.4e-51	183.5	6	61-120:155-214:258-315:348-404:440-497:530-596
1348	fn3	Fibronectin type III domain	4.4e-40	146.6	4	615-704:717-807:819-907:919-1002
1350	serpin	Serpin (serine protease inhibitor)	3.2e-205	695.2	1	46-402
1353	CARD	Caspase recruitment domain	1.3e-32	121.8	1	2-91
1355	ank	Ankyrin repeat	1.1e-45	165.2	6	31-63:64-96:97-129:130-162:163-195:196-228
1356	pkinase	Protein kinase domain	9.6e-64	225.2	1	221-479
1359	tRNA-synt_1	tRNA synthetases class I (I, L, M and V)	1.1e-05	-214.4	1	31-383
1360	MHC_II_beta	Class II histocompatibility antigen, beta	1.7e-41	151.3	1	41-117
1363	ig	Immunoglobulin domain	1.1e-08	42.3	3	114-200:236-294:344-398
1364	Tissue_fac	Tissue factor	0.069	-126.3	1	1-271
1364	fn3	Fibronectin type III domain	0.095	14.9	1	35-125
1365	IL1	Interleukin-1 / 18	7.6e-30	112.6	1	11-155
1366	A2M	Alpha-2-macroglobulin family	1e-210	713.4	1	722-1449
1366	A2M_N	Alpha-2-macroglobulin family N-terminal regi	4.7e-90	312.6	1	1-623
1368	UPAR_LY6	u-PAR/Ly-6 domain	6.8e-37	136.0	1	27-106

490
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
685	Guanylin	Guanylin precursor	0.72	1.2	1	1-27
685	hormone	Somatotropin hormone family	6.7e-18	49.4	1	9-57
685	DUF756	Domain of unknown function (DUF756)	0.4	5.1	1	99-125
686	Guanylin	Guanylin precursor	0.72	1.2	1	1-27
686	hormone	Somatotropin hormone family	3.6e-56	157.9	1	9-151
686	PI3_PI4_kinase	Phosphatidylinositol 3- and 4-kinase	0.97	3.4	1	172-206
688	hormone	Somatotropin hormone family	1.5e-68	192.9	1	9-151
689	serpin	Serpin (serine protease inhibitor)	3.2e-21	71.8	1	49-156
689	serpin	Serpin (serine protease inhibitor)	5.2e-57	193.9	2	160-397
690	PH	PH domain	0.042	8.1	1	1-20
690	efhand	EF hand	9.2e-05	21.0	1	34-62
690	efhand	EF hand	0.0023	15.8	2	70-98
690	PI-PLC-X	Phosphatidylinositol-specific phospho	5.9e-17	60.5	1	187-222
691	Lipase_3	Lipase (class 3)	6.9e-18	63.4	1	366-505
691	Desulfoferrodox	Desulfoferrodoxin	0.9	2.2	1	528-533
692	PH	PH domain	4.7e-05	17.9	1	20-127
692	DUF482	Protein of unknown function, DUF482	0.8	2.7	1	50-67
692	Phage_TAC	Phage tail assembly chaperone	0.21	5.3	1	225-245
692	Glyco_hydro_31	Glycosyl hydrolases family 31	0.8	0.9	1	344-379
692	NHL	NHL repeat	0.25	8.5	1	494-509
692	EspB	Enterobacterial EspB protein	0.27	2.1	1	560-578
694	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) fa	1.6e-55	187.0	1	93-332
694	Ppx-GppA	Ppx/GppA phosphatase family	0.4	3.5	1	249-261
694	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) fa	5.1e-05	15.7	2	430-480
695	7tm_1	7 transmembrane receptor (rhodopsin f	8.1e-28	82.0	1	22-294
695	GSPII_N	Bacterial type II secretion system pr	0.41	3.4	1	110-118
695	GASA	Gibberellin regulated protein	0.72	0.6	1	176-197
696	DUF716	Family of unknown function (DUF716)	0.93	3.4	1	45-73
696	DcuC	C4-dicarboxylate anaerobic carrier	0.4	4.3	1	46-67
696	FLO_LFY	Floricaula / Leafy protein	0.22	2.7	1	146-159
696	lectin_c	Lectin C-type domain	1.9e-07	31.5	1	181-286
696	Rubella_E2	Rubella membrane glycoprotein E2	0.95	1.4	1	284-312
698	CDtoxinC	Cytolethal distending toxin C	0.43	3.9	1	9-33
698	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) fa	1.6e-62	210.7	1	40-275
698	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) fa	0.016	7.2	2	376-393
700	zf-MYND	MYND finger	0.39	5.1	1	173-192
700	Ribosomal_L44	Ribosomal protein L44	0.33	5.8	1	183-208
700	ZZ	Zinc finger, ZZ type	0.0003	17.8	1	184-211
700	PilP	Pilus assembly protein, PilQ	0.028	8.4	1	228-244
700	myb_DNA-binding	Myb-like DNA-binding domain	2.6e-09	37.1	1	231-278
700	RRS1	Ribosome biogenesis regulatory protei	0.85	3.5	1	379-390
701	sigma70_ner	Sigma-70, non-essential region	0.45	3.2	1	616-628
702	zf-AN1	AN1-like Zinc finger	0.032	10.1	1	13-52
702	zf-AN1	AN1-like Zinc finger	9.2e-06	22.6	2	103-135
703	CRAL_TRIO_N	CRAL/TRIO, N-terminus	3.8e-13	44.7	1	3-71
703	DnaJ_C	DnaJ C terminal region	0.054	8.2	1	8-20
703	CRAL_TRIO	CRAL/TRIO domain	1.4e-44	151.9	1	85-244
704	Adrenomedullin	Adrenomedullin	0.82	2.4	1	142-167

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
704	Rhomboid	Rhomboid family	1.6e-14	55.3	1	201-304
705	TRAP_alpha	Translocon-associated protein (TRAP),	0.41	3.2	1	413-434
705	GKAP	Guanylate-kinase-associated protein (2.7e-292	981.2	1	621-979
705	PLRV_ORF5	Potato leaf roll virus readthrough pr	0.13	4.1	1	752-766
705	DUF887	Eukaryotic protein of unknown functio	1	2.6	1	797-815
705	CYTH	CYTH domain	0.26	6.4	1	816-858
705	SeqA	SeqA protein	0.38	3.6	1	824-837
706	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal d	4.5e-36	123.6	1	33-191
706	ABG_transport	AbgT putative transporter family	0.27	2.7	1	196-205
706	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal d	8.3e-14	49.9	1	253-456
706	HS2ST	Heparan sulfate 2-O-sulfotransferase	0.21	4.8	1	309-338
707	Phage_integr_N	Phage integrase, N-terminal SAM-like	0.36	5.2	1	103-121
707	Glyco_transf_8	Glycosyl transferase family 8	0.00044	15.9	1	268-340
708	LIM	LIM domain	9.7e-16	57.8	1	13-69
708	zf-HIT	HIT zinc finger	0.57	6.9	1	55-65
709	DUF572	Family of unknown function (DUF572)	1.9e-204	689.4	1	1-376
710	Collagen	Collagen triple helix repeat (20 copi	1.6e-14	56.8	1	67-126
710	Collagen	Collagen triple helix repeat (20 copi	3.6e-08	32.9	2	127-174
710	Collagen	Collagen triple helix repeat (20 copi	4.1e-07	29.0	3	183-232
710	Collagen	Collagen triple helix repeat (20 copi	0.25	7.3	4	237-254
710	Collagen	Collagen triple helix repeat (20 copi	4.4e-11	43.9	6	293-346
710	Collagen	Collagen triple helix repeat (20 copi	6.4e-07	28.2	7	359-389
710	Collagen	Collagen triple helix repeat (20 copi	0.42	6.4	8	400-418
710	Collagen	Collagen triple helix repeat (20 copi	0.00074	16.8	9	423-448
710	Collagen	Collagen triple helix repeat (20 copi	8.6e-08	31.5	10	451-483
710	Collagen	Collagen triple helix repeat (20 copi	1.1e-11	46.2	11	493-550
710	Collagen	Collagen triple helix repeat (20 copi	6.8e-06	24.4	12	556-593
710	Collagen	Collagen triple helix repeat (20 copi	0.0014	15.7	13	595-622
710	Collagen	Collagen triple helix repeat (20 copi	1.8e-06	26.6	14	624-659
710	Collagen	Collagen triple helix repeat (20 copi	4.1e-12	47.8	15	684-743
710	Collagen	Collagen triple helix repeat (20 copi	2.4e-05	22.3	16	744-774
710	Collagen	Collagen triple helix repeat (20 copi	2e-11	45.2	17	781-829
710	Collagen	Collagen triple helix repeat (20 copi	0.00026	18.5	18	830-859
710	Collagen	Collagen triple helix repeat (20 copi	8.1e-15	57.9	19	860-919
710	Collagen	Collagen triple helix repeat (20 copi	2e-12	48.9	20	920-979
710	Collagen	Collagen triple helix repeat (20 copi	3.5e-06	25.5	21	1000-1031
710	Collagen	Collagen triple helix repeat (20 copi	1.9e-11	45.2	22	1033-1090
710	Collagen	Collagen triple helix repeat (20 copi	6.6e-11	43.2	23	1099-1154
710	Collagen	Collagen triple helix repeat (20 copi	3.9e-13	51.6	24	1155-1214
710	Collagen	Collagen triple helix repeat (20 copi	0.0069	13.1	25	1217-1234
710	Herpes_LP	Herpesvirus leader protein	0.94	2.5	1	1228-1243
710	Collagen	Collagen triple helix repeat (20 copi	0.0001	20.0	26	1238-1269
710	Collagen	Collagen triple helix repeat (20 copi	4e-09	36.5	27	1278-1337

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
710	Collagen	Collagen triple helix repeat (20 copi	1.9e-13	52.8	28	1341-1394
710	Collagen	Collagen triple helix repeat (20 copi	7.1e-06	24.3	29	1401-1434
710	Collagen	Collagen triple helix repeat (20 copi	0.0012	16.0	30	1435-1483
710	C4	C-terminal tandem repeated domain in	2e-69	240.8	1	1489-1596
710	C4	C-terminal tandem repeated domain in	1.3e-77	268.0	2	1597-1711
711	MGAT2	N-acetylglucosaminyltransferase II (M	0.36	0.6	1	61-69
711	ldl_recept_a	Low-density lipoprotein receptor doma	7.7e-15	51.1	1	67-108
711	ldl_recept_a	Low-density lipoprotein receptor doma	4e-10	35.6	2	112-152
711	DUF351	Domain of Unknown Function (DUF351)	0.25	4.8	1	136-144
711	EGF	EGF-like domain	0.00011	19.6	1	157-190
711	EGF	EGF-like domain	0.0004	17.6	2	196-230
711	ldl_recept_b	Low-density lipoprotein receptor repe	7.3e-10	34.9	1	332-373
711	ldl_recept_b	Low-density lipoprotein receptor repe	2.7e-07	26.4	2	375-417
711	ldl_recept_b	Low-density lipoprotein receptor repe	7.6e-08	28.2	3	419-461
711	EGF	EGF-like domain	0.045	10.2	3	512-553
711	ldl_recept_b	Low-density lipoprotein receptor repe	8.3e-10	34.7	4	605-646
711	ldl_recept_b	Low-density lipoprotein receptor repe	8.4e-11	38.1	5	648-692
711	ldl_recept_b	Low-density lipoprotein receptor repe	1.8e-09	33.6	6	694-742
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.00039	15.9	7	744-781
711	EGF	EGF-like domain	0.00036	17.8	4	835-870
711	ldl_recept_a	Low-density lipoprotein receptor doma	6.6e-17	57.9	3	882-920
711	squash	Squash family serine protease inhibit	0.6	2.5	1	892-908
711	ldl_recept_a	Low-density lipoprotein receptor doma	5.8e-15	51.5	4	921-961
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.6e-15	53.3	5	962-1001
711	ldl_recept_a	Low-density lipoprotein receptor doma	2.1e-18	62.8	6	1002-1041
711	DX	DX module	0.78	3.2	1	1016-1047
711	ldl_recept_a	Low-density lipoprotein receptor doma	8.9e-16	54.2	7	1043-1081
711	ldl_recept_a	Low-density lipoprotein receptor doma	2.3e-14	49.5	8	1088-1127
711	ldl_recept_a	Low-density lipoprotein receptor doma	6.8e-11	38.1	9	1130-1170
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.6e-06	23.6	10	1173-1206
711	EGF	EGF-like domain	2.1e-07	29.5	7	1213-1249
711	CBM_14	Chitin binding Peritrophin-A domain	0.1	6.7	1	1235-1255
711	EGF	EGF-like domain	0.099	9.0	8	1255-1289
711	ldl_recept_b	Low-density lipoprotein receptor repe	4.6e-09	32.3	9	1337-1382
711	ldl_recept_b	Low-density lipoprotein receptor repe	6.3e-15	51.8	10	1384-1425
711	ldl_recept_b	Low-density lipoprotein receptor repe	3.3e-11	39.4	11	1427-1472

493
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.094	8.0	12	1474-1515
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.0042	12.5	13	1517-1558
711	EGF	EGF-like domain	0.00016	19.0	9	1568-1606
711	ldl_recept_b	Low-density lipoprotein receptor repe	1.6e-12	43.8	14	1655-1696
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.003	13.0	15	1698-1740
711	ldl_recept_b	Low-density lipoprotein receptor repe	1.7e-07	27.1	16	1742-1780
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.003	13.0	17	1782-1822
711	EGF	EGF-like domain	2.2e-06	25.8	10	1875-1911
711	Keratin	Keratin	0.43	1.6	1	1881-1894
711	DUF244	Uncharacterized protein family (ORF7)	0.77	1.6	1	1934-1952
711	ldl_recept_b	Low-density lipoprotein receptor repe	7.6e-08	28.2	18	1959-2000
711	ldl_recept_b	Low-density lipoprotein receptor repe	2.7e-13	46.3	19	2002-2043
711	ldl_recept_b	Low-density lipoprotein receptor repe	3.1e-11	39.5	20	2045-2087
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.00065	15.2	21	2089-2118
711	EGF	EGF-like domain	8.7e-06	23.6	11	2184-2219
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.49	5.6	22	2318-2365
711	malic_N	Malic enzyme, NAD binding domain	0.26	2.4	1	2340-2362
711	ldl_recept_b	Low-density lipoprotein receptor repe	7.6e-14	48.1	23	2367-2410
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.0026	13.2	24	2412-2440
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.00025	16.6	25	2453-2479
711	EGF	EGF-like domain	0.67	6.0	12	2505-2528
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.5e-14	50.2	11	2545-2586
711	ldl_recept_a	Low-density lipoprotein receptor doma	6.2e-13	44.8	12	2587-2625
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.4e-14	50.2	13	2626-2664
711	ldl_recept_a	Low-density lipoprotein receptor doma	9.4e-11	37.6	14	2682-2713
711	ldl_recept_a	Low-density lipoprotein receptor doma	7.3e-10	34.7	15	2717-2753
711	ldl_recept_a	Low-density lipoprotein receptor doma	5.2e-11	38.5	16	2755-2795
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.8e-17	59.8	17	2796-

494

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						2838
711	ldl_recept_a	Low-density lipoprotein receptor doma	5.8e-14	48.2	18	2840-2879
711	ldl_recept_a	Low-density lipoprotein receptor doma	5.1e-11	38.5	19	2880-2923
711	ldl_recept_a	Low-density lipoprotein receptor doma	5.1e-12	41.8	20	2926-2964
711	EGF	EGF-like domain	0.61	6.1	13	2928-2962
711	dickkopf_N	7/11 2849 2856 .. 47 54	0.32	4.9	8	2935-2942
711	Omega-atracotox	Omega-atracotoxin	0.46	3.7	2	2937-2957
711	EGF	EGF-like domain	3.9e-06	24.9	14	2967-3003
711	TIL	Trypsin Inhibitor like cysteine rich	6.4e-05	16.4	2	2987-3009
711	EGF	EGF-like domain	0.00094	16.3	15	3009-3034
711	ldl_recept_b	Low-density lipoprotein receptor repe	8.1e-09	31.5	26	3092-3134
711	ldl_recept_b	Low-density lipoprotein receptor repe	4.1e-07	25.8	27	3136-3177
711	ldl_recept_b	Low-density lipoprotein receptor repe	1.1e-08	31.0	28	3179-3221
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.078	8.3	29	3223-3251
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.0013	14.2	30	3262-3289
711	EGF	EGF-like domain	1.6e-06	26.3	16	3314-3350
711	TNFR_c6	1/3 69 84 .. 1 18	0.42	6.2	2	3337-3352
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.9e-12	43.2	21	3352-3391
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.4e-12	43.7	22	3392-3430
711	ldl_recept_a	Low-density lipoprotein receptor doma	3.9e-12	42.2	23	3431-3470
711	ldl_recept_a	Low-density lipoprotein receptor doma	3.5e-17	58.8	24	3471-3510
711	SAPA	Saposin A-type domain	0.039	6.0	1	3479-3492
711	Sar8_2	Sar8.2 family	0.12	6.9	1	3480-3500
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.4e-19	66.7	25	3511-3549
711	ldl_recept_a	Low-density lipoprotein receptor doma	8.3e-13	44.4	26	3550-3588
711	EGF	EGF-like domain	0.54	6.3	17	3552-3586
711	ldl_recept_a	Low-density lipoprotein receptor doma	1.3e-14	50.4	27	3590-3626
711	dickkopf_N	7/11 2849 2856 .. 47 54	0.057	7.2	10	3596-3604

495
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
711	ldl_recept_a	Low-density lipoprotein receptor doma	6e-14	48.1	28	3629-3666
711	Herpes_PAP	Herpesvirus polymerase accessory prot	0.41	2.1	1	3637-3650
711	ldl_recept_a	Low-density lipoprotein receptor doma	2e-19	66.2	29	3667-3706
711	ldl_recept_a	Low-density lipoprotein receptor doma	2.7e-11	39.4	30	3709-3749
711	ldl_recept_a	Low-density lipoprotein receptor doma	5.7e-07	25.1	31	3758-3790
711	ldl_recept_a	Low-density lipoprotein receptor doma	5.3e-17	58.2	32	3797-3835
711	EGF	19/28 3669 3704 .. 1 46	0.00016	19.1	20	3842-3879
711	S_locus_glycop	S-locus glycoprotein family	0.94	5.0	1	3849-3870
711	TIL	Trypsin Inhibitor like cysteine rich	0.051	7.3	3	3864-3885
711	laminin_EGF	Laminin EGF-like (Domains III and V)	0.23	6.5	2	3865-3879
711	EGF	19/28 3669 3704 .. 1 46	0.0054	13.5	21	3885-3914
711	Activin_recp	Activin types I and II receptor domai	0.48	3.4	1	3891-3921
711	NHL	2/5 681 694 .. 1 14	0.06	10.7	4	4005-4031
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.14	7.5	31	4008-4016
711	ldl_recept_b	Low-density lipoprotein receptor repe	0.11	7.8	32	4018-4026
711	ldl_recept_b	33/35 4040 4074 .. 9 47	1.9e-10	36.9	34	4076-4118
711	ldl_recept_b	33/35 4040 4074 .. 9 47	0.019	10.3	35	4120-4163
711	EGF	19/28 3669 3704 .. 1 46	0.77	5.8	22	4213-4236
711	EB	EB module	0.15	6.2	3	4229-4244
711	EGF	19/28 3669 3704 .. 1 46	0.00038	17.7	23	4254-4285
711	EGF	19/28 3669 3704 .. 1 46	8.8e-08	30.8	24	4290-4321
711	EGF	19/28 3669 3704 .. 1 46	7.4e-08	31.1	25	4326-4357
711	EGF	27/28 4398 4428 .. 1 46	0.0014	15.6	28	4431-4463
711	Coagulin	Coagulin	0.52	3.4	1	4447-4454
711	Herpes_glycop_D	Herpesvirus glycoprotein D	0.39	4.3	2	4483-4519
712	MGAT2	N-acetylglucosaminyltransferase II (M	0.36	0.6	1	61-69
712	ldl_recept_a	Low-density lipoprotein receptor doma	1e-14	50.7	1	67-108
712	ldl_recept_a	Low-density lipoprotein receptor doma	4e-10	35.6	2	112-152
712	DUF351	Domain of Unknown Function (DUF351)	0.25	4.8	1	136-144

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
712	EGF	EGF-like domain	0.072	9.5	2	157-181
714	cadherin	Cadherin domain	0.085	8.0	1	47-65
714	cadherin	Cadherin domain	0.00072	15.2	2	69-126
714	cadherin	Cadherin domain	8.4e-17	60.3	3	140-241
714	cadherin	Cadherin domain	1.4e-29	104.9	4	255-344
714	cadherin	Cadherin domain	7.9e-25	88.3	5	363-466
714	cadherin	Cadherin domain	2e-26	93.9	6	480-573
714	cadherin	Cadherin domain	3.2e-28	100.1	7	588-680
714	Rad21_Rec8	Conserved region of Rad21 / Rec8 like	0.83	5.2	1	652-662
714	cadherin	Cadherin domain	3.9e-28	99.8	8	694-784
714	SCPU	Spore Coat Protein U domain	0.47	5.3	1	701-714
714	cadherin	Cadherin domain	5.7e-20	71.3	9	798-884
714	cadherin	Cadherin domain	7.6e-20	70.9	10	898-987
714	cadherin	Cadherin domain	9.5e-28	98.5	11	1001-1091
714	cadherin	Cadherin domain	5.1e-16	57.6	12	1105-1201
714	cadherin	Cadherin domain	1.4e-28	101.4	13	1215-1306
714	Propep_M14	Carboxypeptidase activation peptide	0.41	5.5	2	1228-1239
714	cadherin	Cadherin domain	2.2e-29	104.2	14	1320-1411
714	cadherin	Cadherin domain	7.2e-21	74.5	15	1425-1520
714	Baculo_helicase	Baculovirus DNA helicase	0.61	1.4	1	1521-1531
714	cadherin	Cadherin domain	4.5e-16	57.7	16	1541-1622
714	cadherin	Cadherin domain	0.00017	17.4	17	1634-1700
715	cadherin	Cadherin domain	0.085	8.0	1	47-65
715	cadherin	Cadherin domain	0.00072	15.2	2	69-126
715	cadherin	Cadherin domain	8.4e-17	60.3	3	140-241
715	cadherin	Cadherin domain	1.4e-29	104.9	4	255-344
715	cadherin	Cadherin domain	6.1e-25	88.7	5	363-466
715	cadherin	Cadherin domain	2e-26	93.9	6	480-573
715	cadherin	Cadherin domain	3.2e-28	100.1	7	588-680
715	Rad21_Rec8	Conserved region of Rad21 / Rec8 like	0.83	5.2	1	652-662
715	cadherin	Cadherin domain	3.9e-28	99.8	8	694-784
715	SCPU	Spore Coat Protein U domain	0.47	5.3	1	701-714
715	cadherin	Cadherin domain	5.7e-20	71.3	9	798-884
715	cadherin	Cadherin domain	7.6e-20	70.9	10	898-987
715	cadherin	Cadherin domain	9.5e-28	98.5	11	1001-1091
715	cadherin	Cadherin domain	5.1e-16	57.6	12	1105-1201
715	cadherin	Cadherin domain	1.4e-28	101.4	13	1215-1306
715	Propep_M14	Carboxypeptidase activation peptide	0.41	5.5	2	1228-1239
715	cadherin	Cadherin domain	2.2e-29	104.2	14	1320-1411
715	cadherin	Cadherin domain	7.2e-21	74.5	15	1425-

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						1520
715	Baculo_helicase	Baculovirus DNA helicase	0.61	1.4	1	1521-1531
715	cadherin	Cadherin domain	4.5e-16	57.7	16	1541-1622
715	cadherin	Cadherin domain	3.5e-05	19.8	17	1634-1728
716	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-te	0.5	1.1	1	310-346
716	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-te	0.0014	8.6	2	516-589
716	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-te	5.3e-08	21.7	3	618-652
716	Peptidase_S9	Prolyl oligopeptidase family	3.9e-11	36.8	1	664-736
716	Methyltransf_6	Demethylmenaquinone methyltransferase	0.54	3.9	1	675-688
716	Esterase	Putative esterase	0.062	6.6	1	710-753
717	zf-C2H2	Zinc finger, C2H2 type	0.015	14.8	1	32-54
717	zf-C2H2	Zinc finger, C2H2 type	0.0014	18.9	2	60-82
717	Apocytochr_F_C	Apocytochrome F, C-terminal	1	3.2	1	103-110
717	TFIIS	Transcription factor S-II (TFIIS)	0.2	7.1	1	154-164
717	zf-C2H2	Zinc finger, C2H2 type	3.7e-08	37.3	3	154-176
717	XPA_N	XPA protein N-terminal	0.3	6.5	2	179-191
717	zf-C2H2	Zinc finger, C2H2 type	8.5e-06	27.9	4	182-204
717	zf-C2H2	Zinc finger, C2H2 type	6.4e-08	36.5	5	210-232
717	TFIIS	3/8 210 220 .. 29 39	1	4.7	4	238-248
717	zf-C2H2	Zinc finger, C2H2 type	1.6e-06	30.8	6	238-260
717	XPA_N	XPA protein N-terminal	1	4.6	4	263-275
717	zf-C2H2	Zinc finger, C2H2 type	1.4e-05	27.0	7	266-288
717	zf-C2H2	Zinc finger, C2H2 type	2.6e-05	25.9	8	294-316
717	TFIIS	5/8 266 276 .. 29 39	0.2	7.1	7	322-332
717	zf-C2H2	Zinc finger, C2H2 type	6.9e-06	28.3	9	322-344
717	XPA_N	XPA protein N-terminal	0.38	6.2	6	347-359
717	TFIIS	5/8 266 276 .. 29 39	0.14	7.7	8	350-360
717	zf-C2H2	Zinc finger, C2H2 type	1e-07	35.7	10	350-372
719	Phytoreo_Pns	Phytoreovirus nonstructural protein P	0.75	2.1	1	74-88
719	malic	Malic enzyme, N-terminal domain	0.39	3.5	1	117-131
719	AlpA	Prophage CP4-57 regulatory protein (A	0.95	4.3	1	258-266
719	DUF298	Domain of unknown function (DUF298)	0.42	5.1	1	308-337
719	DUF827	Plant protein of unknown function (DU	0.029	7.3	1	363-387
719	DUF496	Protein of unknown function (DUF496)	0.49	5.1	1	389-409
719	K-box	K-box region	0.37	5.2	1	392-406
719	TFIIE_alpha	TFIIE alpha subunit	0.14	5.9	1	394-416
719	Mlp	Mlp lipoprotein family	0.95	2.4	1	398-451
719	Ribosomal_S20p	Ribosomal protein S20	0.38	5.2	1	433-447
719	Phage_B	Scaffold protein B	0.47	1.7	1	504-518
720	ig	Immunoglobulin domain	0.07	9.9	1	17-34
720	ig	Immunoglobulin domain	5.1e-11	44.1	2	68-128
720	ig	Immunoglobulin domain	1.1e-11	46.7	3	163-223
720	ig	Immunoglobulin domain	9.6e-07	28.1	4	259-317
720	AstA	Arginine N-succinyltransferase beta s	0.92	2.5	1	294-305
720	ig	Immunoglobulin domain	2.1e-09	38.1	5	352-410
720	ig	Immunoglobulin domain	1.5e-10	42.3	6	445-503
720	RTC	RNA 3'-terminal phosphate cyclase	0.7	3.3	1	474-491
720	ig	Immunoglobulin domain	8.1e-08	32.1	7	538-596
720	ig	Immunoglobulin domain	1.3e-07	31.3	8	629-687

498

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
720	ig	Immunoglobulin domain	8.1e-09	35.9	9	720-780
720	ig	Immunoglobulin domain	3.7e-09	37.2	10	813-871
720	ig	Immunoglobulin domain	7.4e-09	36.0	11	904-962
720	ig	Immunoglobulin domain	1.9e-11	45.7	12	995-1052
720	ig	Immunoglobulin domain	1.3e-07	31.4	13	1085-1143
720	ig	Immunoglobulin domain	1.3e-11	46.4	14	1176-1232
720	ig	Immunoglobulin domain	3.6e-10	40.9	15	1266-1323
720	Marek_A	Marek's disease glycoprotein A	0.84	1.1	1	1333-1356
720	RNA_pol_Rpb2_1	RNA polymerase beta subunit	0.35	1.6	1	1352-1864
720	ig	Immunoglobulin domain	6.4e-10	40.0	16	1356-1413
720	tsp_1	Thrombospondin type 1 domain	1.2e-19	67.2	1	1435-1485
720	tsp_1	Thrombospondin type 1 domain	6.4e-17	58.1	2	1492-1542
720	tsp_1	Thrombospondin type 1 domain	3.5e-15	52.3	3	1549-1599
720	tsp_1	Thrombospondin type 1 domain	2.2e-17	59.7	4	1606-1656
720	tsp_1	Thrombospondin type 1 domain	8.2e-12	41.1	5	1663-1713
720	VOMI	Vitelline membrane outer layer protei	0.37	3.6	1	1714-1728
720	tsp_1	Thrombospondin type 1 domain	7e-16	54.7	6	1720-1770
720	EGF	EGF-like domain	0.95	5.4	1	1993-2007
720	EGF	EGF-like domain	9.3e-08	30.7	2	2013-2047
720	granulin	Granulin	0.44	4.7	1	2034-2049
720	EGF	EGF-like domain	0.015	11.9	3	2053-2092
720	EGF	EGF-like domain	2.8e-05	21.8	4	2098-2130
720	TIL	1/7 1698 1715 .. 1 16	0.0012	12.5	3	2117-2136
720	EGF	EGF-like domain	0.17	8.2	5	2136-2157
720	EGF	EGF-like domain	2.4e-06	25.7	6	2178-2215
720	EGF	EGF-like domain	5.7e-10	38.7	7	2221-2256
720	Ribosomal_L34	Ribosomal protein L34	0.33	5.5	1	2280-2323
720	TIL	4/7 2168 2178 .. 57 68	0.022	8.5	6	2320-2338
720	EGF	EGF-like domain	1.9e-09	36.8	8	2338-2372

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
720	toxin_2	Scorpion short toxin	0.84	3.4	2	2338-2353
720	toxin_5	Scorpion short toxin	0.73	3.4	1	2354-2359
720	TIL	4/7 2168 2178 .. 57 68	0.023	8.4	7	2357-2378
720	squash	Squash family serine protease inhibit	0.44	2.8	2	2358-2386
720	fn2	Fibronectin type II domain	0.8	3.1	1	2407-2418
721	SAP	SAP domain	2.4e-10	40.8	1	3-37
721	SPRY	SPRY domain	1.8e-30	107.5	1	289-418
721	SRP54	SRP54-type protein, GTPase domain	0.0091	11.6	1	451-466
721	NACHT	NACHT domain	0.18	5.5	1	453-469
721	SKI	Shikimate kinase	0.33	4.9	1	453-466
721	Zot	Zonular occludens toxin (Zot)	0.22	5.5	1	453-466
721	AAA	ATPase family associated with various	0.098	5.8	1	454-466
721	tRNA_synt_1c_R2	Glutaminyl-tRNA synthetase, non-speci	0.79	3.9	1	580-616
722	CheB_methylest	CheB methylesterase	1	2.7	1	74-92
722	DUF258	Protein of unknown function, DUF258	0.0014	13.8	1	509-532
722	ABC_tran	ABC transporter	7.4e-59	198.4	1	510-692
722	NACHT	NACHT domain	0.2	5.3	1	511-527
722	SMC_N	RecF/RecN/SMC N terminal domain	0.47	3.9	1	511-524
722	Zot	Zonular occludens toxin (Zot)	0.28	5.1	1	511-524
722	RHD3	Root hair defective 3 GTP-binding pro	0.67	1.2	1	516-530
722	Pox_D2	Pox virus D2 protein	0.86	1.4	1	604-617
722	tail_comp_S	Phage virion morphogenesis family	0.061	7.3	1	606-619
722	DUF333	Domain of unknown function (DUF333)	0.3	5.7	1	818-846
722	ABC_tran	ABC transporter	1.1e-47	160.9	2	1322-1506
722	SufE	Fe-S metabolism associated domain	0.28	6.2	1	1544-1563
723	BEX	Brain expressed X-linked like family	0.88	2.2	1	133-160
723	CytoC_RC	Photosynthetic reaction centre cytoch	1	1.4	1	215-231
723	Ski_Sno	SKI/SNO/DAC family	0.51	4.5	1	656-672
724	HpaB	4-hydroxyphenylacetate 3-hydroxylase	0.97	2.5	1	4-14
724	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal do	6.7e-50	175.9	1	50-201
725	C_tripleX	Cysteine rich repeat	2e-05	17.8	1	59-76
725	Bowman-Birk_leg	Bowman-Birk serine protease inhibitor	1	4.0	1	68-83
725	laminin_EGF	Laminin EGF-like (Domains III and V)	0.32	6.1	1	80-93
725	EGF	EGF-like domain	8.7e-06	23.6	2	98-126
725	TIL	Trypsin Inhibitor like cysteine rich	0.0035	11.0	1	117-138
725	EGF	EGF-like domain	7.5e-05	20.2	3	138-172
725	TIL	Trypsin Inhibitor like cysteine rich	0.26	5.1	2	151-178
725	toxin_5	Scorpion short toxin	0.34	4.4	1	153-158
725	EGF	EGF-like domain	4.4e-05	21.1	4	178-211
725	EGF	EGF-like domain	9.7e-09	34.3	5	223-258
725	MAM	MAM domain	3.5e-41	147.0	1	402-546
726	DUF626	Protein of unknown function (DUF626)	0.22	5.8	1	30-64
726	VSP	Giardia variant-specific surface prot	1	1.8	1	106-131

500

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
726	zf-B_box	B-box zinc finger	5.9e-08	32.7	1	106-139
726	Prefoldin	Prefoldin subunit	0.42	5.7	1	222-248
726	Filamin	Filamin/ABP280 repeat	2.3e-21	74.7	1	313-402
726	NHL	NHL repeat	4.4e-10	40.3	1	431-458
726	Glyoxalase	Glyoxalase/Bleomycin resistance prote	0.78	3.6	1	476-504
726	NHL	NHL repeat	2.4e-10	41.2	2	478-505
726	NHL	NHL repeat	1.1e-10	42.4	3	525-552
726	NHL	NHL repeat	2.5e-09	37.6	4	572-599
726	NHL	NHL repeat	7.8e-11	43.0	5	619-646
726	NHL	NHL repeat	3.8e-08	33.2	6	666-693
727	PaREP1	Archaeal PaREP1 protein	0.38	5.3	1	111-127
727	FCH	Fes/CIP4 homology domain	0.026	10.3	1	281-321
727	DAG_PE-bind	Phorbol esters/diacylglycerol binding	2.8e-05	21.7	1	709-747
727	RhoGAP	RhoGAP domain	3.9e-68	231.7	1	775-947
727	Terpene_synth_C	Terpene synthase family, metal bindin	0.84	2.7	1	778-812
727	NnrS	NnrS protein	1	1.8	1	934-943
728	DUF727	Protein of unknown function (DUF727)	0.83	4.2	1	115-129
728	CN_hydrolase	Carbon-nitrogen hydrolase	4e-09	33.8	2	120-216
729	Pep_M12B_prop ep	Reprolysin family propeptide	3.3e-14	44.8	1	93-223
729	Reprolysin	Reprolysin (M12B) family zinc metallo	0.00037	16.1	1	274-296
729	PsaL	Photosystem I reaction centre subunit	0.99	3.2	1	302-317
729	Reprolysin	Reprolysin (M12B) family zinc metallo	8.5e-17	62.5	2	340-480
729	Fragilysin	Fragilysin metalloproteinase (M10C) en	0.46	3.1	1	412-430
729	dickkopf_N	Dickkopf N-terminal cysteine-rich reg	0.0036	10.8	1	534-560
729	Stig1	Stigma-specific protein, Stig1	0.11	4.5	1	544-558
729	EB	EB module	0.8	3.9	1	546-558
729	tsp_1	Thrombospondin type 1 domain	7.1e-09	31.3	1	570-623
729	zf-A20	A20-like zinc finger	0.39	8.6	1	702-717
729	ADAM_spacer1	ADAM-TS Spacer 1	3.8e-49	173.5	1	734-852
729	Herpes_VP19C	Herpesvirus capsid shell protein VP19	0.95	3.6	1	860-871
729	tsp_1	2/12 866 875 .. 4 13	0.048	8.5	3	985-1002
729	tsp_1	2/12 866 875 .. 4 13	0.067	8.1	4	1037-1089
729	tsp_1	2/12 866 875 .. 4 13	1.2e-05	20.6	5	1092-1115
729	PTN_MK_N	PTN/MK heparin-binding protein family	0.44	4.2	1	1165-1184
729	tsp_1	2/12 866 875 .. 4 13	7.6e-07	24.5	6	1165-1190
729	tsp_1	2/12 866 875 .. 4 13	1.4e-06	23.7	7	1228-1276
729	tsp_1	2/12 866 875 .. 4 13	4.6e-07	25.3	8	1313-1364
729	tsp_1	2/12 866 875 .. 4 13	0.00029	15.9	9	1372-1420
729	tsp_1	2/12 866 875 .. 4 13	1.7e-07	26.7	10	1426-1479
729	tsp_1	2/12 866 875 .. 4 13	4.7e-05	18.6	11	1485-1506
729	tsp_1	2/12 866 875 .. 4 13	0.00073	14.6	12	1543-1593
730	Adeno_Penton_B	Adenovirus penton base protein	0.39	1.6	1	178-193

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
731	ig	Immunoglobulin domain	0.19	8.3	1	6-99
731	DUF390	Protein of unknown function (DUF390)	0.73	0.9	1	83-95
731	ig	Immunoglobulin domain	9.7e-05	20.6	2	146-235
731	ig	Immunoglobulin domain	0.00014	20.0	3	282-373
732	ig	Immunoglobulin domain	0.0045	14.4	1	42-129
732	ig	Immunoglobulin domain	0.19	8.3	2	179-272
732	DUF390	Protein of unknown function (DUF390)	0.73	0.9	1	256-268
732	ig	Immunoglobulin domain	9.7e-05	20.6	3	319-408
732	ig	Immunoglobulin domain	0.00014	20.0	4	455-546
733	ig	Immunoglobulin domain	0.0045	14.4	1	42-129
734	ig	Immunoglobulin domain	0.0018	15.8	1	42-126
734	DUF390	Protein of unknown function (DUF390)	0.73	0.9	1	110-122
735	RhoGEF	RhoGEF domain	8.2e-08	27.0	1	165-225
735	FA_hydroxylase	Fatty acid hydroxylase	0.6	3.7	1	221-233
735	RhoGEF	RhoGEF domain	7.5e-09	30.5	2	257-329
736	HEM4	Uroporphyrinogen-III synthase HemD	0.98	3.1	1	549-581
736	DUF178	Uncharacterized ACR, COG1427	0.11	6.0	1	604-622
737	rrm	RNA recognition motif. (a.k.a. RRM, R	2.5e-07	28.2	1	78-142
737	Smg4_UPF3	Smg-4/UPF3 family	0.042	8.7	1	143-173
737	rrm	RNA recognition motif. (a.k.a. RRM, R	9.7e-16	58.1	2	151-222
737	fer4_NifH	4Fe-4S iron sulfur cluster binding pr	1	2.4	1	160-176
737	rrm	RNA recognition motif. (a.k.a. RRM, R	3.6e-06	24.1	3	274-311
738	Adeno_E4_34	Adenovirus early E4 34 kDa protein co	0.45	4.4	1	5-22
739	ribonuc_red_sm	Ribonucleotide reductase, small chain	0.29	3.7	1	244-265
740	Sua5_yciO_yrdC	yrdC domain	0.99	3.3	1	38-53
740	F-box	F-box domain	0.095	9.0	1	134-175
740	DUF469	Protein with unknown function (DUF469	0.38	4.7	1	354-371
741	OmpH	Outer membrane protein (OmpH-like)	0.14	6.9	1	81-150
741	Herpes_BLRF2	Herpesvirus BLRF2 protein	0.12	7.3	1	256-277
741	UIM	Ubiquitin interaction motif	0.34	8.8	1	293-310
741	DUF260	Protein of unknown function DUF260	0.26	4.8	1	330-350
741	TelA	Toxic anion resistance protein (TelA)	0.34	4.5	1	348-368
741	Pox_A_type_inc	1/5 216 235 .. 1 23	0.6	6.3	2	358-377
741	PspA_IM30	PspA/IM30 family	0.34	5.2	1	364-399
741	M	1/5 272 292 .. 1 21	0.46	8.0	3	534-554
741	Coprinus_mating	Coprinus cinereus mating-type protein	0.65	1.6	1	698-729
741	Ribosomal_L29e	Ribosomal L29e protein family	0.3	5.8	1	717-755
741	Phage_portal_2	Phage portal protein, lambda family	0.75	2.2	1	799-816
741	Dishevelled	Dishevelled specific domain	0.22	4.9	1	903-922
741	SlyX	SlyX	0.69	1.3	1	945-954
742	cadherin	Cadherin domain	0.13	7.4	1	30-96
742	cadherin	Cadherin domain	8.4e-13	46.4	2	147-243
742	cadherin	Cadherin domain	7.1e-25	88.5	3	257-349
742	He_PIG	Putative Ig domain	0.4	5.5	1	262-279
742	cadherin	Cadherin domain	0.049	8.8	4	369-399
742	cadherin	Cadherin domain	2.3e-05	20.4	5	427-460
742	cadherin	Cadherin domain	5.6e-21	74.9	6	474-563
742	cadherin	Cadherin domain	1.9e-25	90.4	7	577-666
742	cadherin	Cadherin domain	4.5e-09	33.4	8	693-737
743	PGM_PMM_I	Phosphoglucomutase/phosphomannomutase	1.6e-15	57.2	1	1-47
743	PGM_PMM	Phosphoglucomutase/phosphomannomutase	0.041	9.3	1	388-430

502

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
743	Cor1	Cor1/Xlr/Xmr conserved region	0.73	4.1	1	425-435
744	MACPF	MAC/Perforin domain	0.00017	15.5	1	138-170
744	Keratin_matx	Keratin, high-sulphur matrix protein	0.19	7.6	1	451-482
744	Nol1_Nop2_Sun	NOL1/NOP2/sun family	0.29	4.1	1	602-622
745	Remorin_C	Remorin, C-terminal region	0.19	6.6	1	120-133
745	zf-C2H2	Zinc finger, C2H2 type	0.00033	21.5	1	130-152
745	TFIIS	Transcription factor S-II (TFIIS)	0.3	6.5	1	158-168
745	zf-C2H2	Zinc finger, C2H2 type	4.3e-07	33.1	2	158-180
745	XPA_N	XPA protein N-terminal	0.72	5.2	3	183-195
745	TFIIS	Transcription factor S-II (TFIIS)	0.069	8.7	2	186-196
745	zf-C2H2	Zinc finger, C2H2 type	3.9e-07	33.3	3	186-208
745	XPA_N	XPA protein N-terminal	0.21	7.1	4	211-223
745	zf-C2H2	Zinc finger, C2H2 type	9.4e-08	35.8	4	214-236
745	zf-C2H2	Zinc finger, C2H2 type	4.8e-07	32.9	5	242-264
745	XPA_N	XPA protein N-terminal	0.13	7.8	6	267-279
745	TFIIS	Transcription factor S-II (TFIIS)	0.28	6.6	5	270-280
745	zf-C2H2	Zinc finger, C2H2 type	3.1e-07	33.7	6	270-292
745	XPA_N	XPA protein N-terminal	0.13	7.8	7	295-307
745	TFIIS	Transcription factor S-II (TFIIS)	0.54	5.6	6	298-308
745	zf-C2H2	Zinc finger, C2H2 type	3.1e-06	29.7	7	298-320
745	XPA_N	XPA protein N-terminal	0.74	5.2	8	323-335
745	TFIIS	Transcription factor S-II (TFIIS)	0.073	8.6	7	325-336
745	zf-C2H2	Zinc finger, C2H2 type	3.3e-07	33.6	8	326-348
745	XPA_N	XPA protein N-terminal	0.72	5.2	9	351-363
745	TFIIS	Transcription factor S-II (TFIIS)	0.51	5.7	8	354-364
745	zf-C2H2	Zinc finger, C2H2 type	7.5e-07	32.2	9	354-376
745	XPA_N	XPA protein N-terminal	0.13	7.8	10	379-391
745	TFIIS	Transcription factor S-II (TFIIS)	0.28	6.6	9	382-392
745	zf-C2H2	Zinc finger, C2H2 type	4.4e-06	29.1	10	382-404
745	XPA_N	XPA protein N-terminal	0.13	7.8	11	407-419
745	TFIIS	Transcription factor S-II (TFIIS)	0.28	6.6	10	410-420
745	zf-C2H2	Zinc finger, C2H2 type	2.7e-07	33.9	11	410-432
745	zf-C2H2	Zinc finger, C2H2 type	0.0011	19.4	12	440-460
745	XPA_N	XPA protein N-terminal	0.67	5.3	12	485-497
745	zf-C2H2	13/16 466 481 .. 1 17	3.9e-06	29.3	14	488-510
745	TFIIS	Transcription factor S-II (TFIIS)	0.0051	12.6	12	515-526
745	zf-C2H2	13/16 466 481 .. 1 17	1.3e-05	27.2	15	516-538
745	zf-BED	BED zinc finger	0.71	4.6	3	517-539
745	XPA_N	XPA protein N-terminal	0.092	8.3	14	541-553
745	TFIIS	Transcription factor S-II (TFIIS)	0.28	6.6	13	544-554
745	zf-C2H2	13/16 466 481 .. 1 17	0.00057	20.5	16	544-565
746	KRAB	KRAB box	6.9e-24	88.6	1	35-75
746	ROS_MUCR	ROS/MUCR transcriptional regulator pr	0.33	3.9	1	81-104
746	Remorin_C	Remorin, C-terminal region	0.19	6.6	1	195-208
746	zf-C2H2	Zinc finger, C2H2 type	0.00033	21.5	1	205-227
746	TFIIS	Transcription factor S-II (TFIIS)	0.3	6.5	1	233-243
746	zf-C2H2	Zinc finger, C2H2 type	4.3e-07	33.1	2	233-255
746	XPA_N	XPA protein N-terminal	0.72	5.2	3	258-270
746	TFIIS	Transcription factor S-II (TFIIS)	0.069	8.7	2	261-271
746	zf-C2H2	Zinc finger, C2H2 type	3.9e-07	33.3	3	261-283
746	XPA_N	XPA protein N-terminal	0.21	7.1	4	286-298
746	zf-C2H2	Zinc finger, C2H2 type	9.4e-08	35.8	4	289-311
746	zf-C2H2	Zinc finger, C2H2 type	4.8e-07	32.9	5	317-339

503
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
746	XPA_N	XPA protein N-terminal	0.13	7.8	6	342-354
746	TFIIS	Transcription factor S-II (TFIIS)	0.28	6.6	5	345-355
746	zf-C2H2	Zinc finger, C2H2 type	3.1e-07	33.7	6	345-367
746	XPA_N	XPA protein N-terminal	0.13	7.8	7	370-382
746	TFIIS	Transcription factor S-II (TFIIS)	0.54	5.6	6	373-383
746	zf-C2H2	Zinc finger, C2H2 type	3.1e-06	29.7	7	373-395
746	XPA_N	XPA protein N-terminal	0.74	5.2	8	398-410
746	TFIIS	Transcription factor S-II (TFIIS)	0.073	8.6	7	400-411
746	zf-C2H2	Zinc finger, C2H2 type	3.3e-07	33.6	8	401-423
746	XPA_N	XPA protein N-terminal	0.72	5.2	9	426-438
746	TFIIS	Transcription factor S-II (TFIIS)	0.51	5.7	8	429-439
746	zf-C2H2	Zinc finger, C2H2 type	7.5e-07	32.2	9	429-451
746	XPA_N	XPA protein N-terminal	0.13	7.8	10	454-466
746	TFIIS	Transcription factor S-II (TFIIS)	0.28	6.6	9	457-467
746	zf-C2H2	Zinc finger, C2H2 type	4.4e-06	29.1	10	457-479
746	XPA_N	XPA protein N-terminal	0.13	7.8	11	482-494
746	TFIIS	Transcription factor S-II (TFIIS)	0.28	6.6	10	485-495
746	zf-C2H2	Zinc finger, C2H2 type	2.7e-07	33.9	11	485-507
746	zf-C2H2	Zinc finger, C2H2 type	0.0011	19.4	12	515-535
747	EMP24_GP25L	emp24/gp25L/p24 family	4.9e-80	276.1	1	5-201
748	acid_phosphat	Histidine acid phosphatase	7.9e-159	537.8	1	31-371
749	C_tripleX	Cysteine rich repeat	0.92	4.2	1	52-67
749	ApoC-I	Apolipoprotein C-I (ApoC-1)	0.83	3.7	1	196-260
749	PH	PH domain	1.5e-20	69.0	1	393-487
749	ArfGap	Putative GTPase activating protein fo	2.1e-60	210.7	1	527-649
749	ank	1/4 797 823 .. 7 33	1.5e-08	33.7	2	826-858
749	ank	1/4 797 823 .. 7 33	0.0001	20.0	3	859-891
751	DUF369	Domain of unknown function (DUF369)	0.17	5.8	1	275-288
751	KRAB	KRAB box	1.1e-20	77.0	1	342-382
751	zf-C2H2	Zinc finger, C2H2 type	7.8e-06	28.0	1	603-625
751	TFIIS	Transcription factor S-II (TFIIS)	0.78	5.1	1	604-613
751	zf-C2H2	Zinc finger, C2H2 type	1.6e-05	26.8	2	631-653
751	zf-C2H2	Zinc finger, C2H2 type	3.7e-07	33.4	3	693-715
751	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.54	2.9	1	708-726
751	TFIIS	Transcription factor S-II (TFIIS)	0.63	5.4	3	721-731
751	zf-C2H2	Zinc finger, C2H2 type	1.3e-05	27.2	4	721-743
751	zf-C2H2	Zinc finger, C2H2 type	3.4e-08	37.4	5	751-773
751	zf-BED	BED zinc finger	0.31	5.8	1	752-774
751	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.032	6.3	2	766-784
751	zf-C2H2	Zinc finger, C2H2 type	5.7e-06	28.6	6	779-801
752	Vps16_N	Vps16, N-terminal region	2.3e-273	918.3	1	1-420
752	Ribosomal_L36	Ribosomal protein L36	0.6	5.0	1	245-281
752	Fumerase	Fumarate hydratase (Fumerase)	0.71	3.4	1	376-402
752	Peptidase_M16_C	Peptidase M16 inactive domain	0.29	5.2	1	492-510
752	Vps16_C	Vps16, C-terminal region	2.4e-15	57.9	1	517-548
752	Vps16_C	Vps16, C-terminal region	4.6e-128	435.6	2	554-762
753	LRRNT	Leucine rich repeat N-terminal domain	0.0011	14.5	1	30-59
753	XG_FTase	Xyloglucan fucosyltransferase	0.53	2.0	1	37-48
753	LRR	Leucine Rich Repeat	0.36	6.7	1	61-82

504

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
753	LRR	Leucine Rich Repeat	0.0014	14.8	2	83-106
753	LRR	Leucine Rich Repeat	5.7e-05	19.5	3	107-131
753	LRR	Leucine Rich Repeat	2.7e-05	20.6	4	132-155
753	LRR	Leucine Rich Repeat	0.001	15.3	5	156-179
753	LRR	Leucine Rich Repeat	0.0036	13.4	6	180-203
753	LRR	Leucine Rich Repeat	0.0016	14.6	7	204-227
753	LRR	Leucine Rich Repeat	0.00015	18.1	8	228-251
753	LRRCT	Leucine rich repeat C-terminal domain	9.7e-12	37.1	1	261-311
754	A2M_N	Alpha-2-macroglobulin family N-termin	4.5e-91	312.7	1	6-613
754	Big_1	Bacterial Ig-like domain (group 1)	0.62	3.9	1	382-403
754	A2M	Alpha-2-macroglobulin family	6.2e-64	214.2	1	721-949
754	A2M	Alpha-2-macroglobulin family	6.2e-132	444.2	2	983-1469
754	Pox_D2	Pox virus D2 protein	0.18	3.4	1	1446-1461
755	DUF904	Protein of unknown function (DUF904)	0.21	6.7	1	116-125
755	DUF536	Protein of unknown function, DUF536	0.47	6.4	1	162-192
755	Syntaxin	Syntaxin	0.11	7.9	1	163-197
755	fibrinogen_C	Fibrinogen beta and gamma chains, C-t	1.7e-09	32.1	1	242-275
755	fibrinogen_C	Fibrinogen beta and gamma chains, C-t	1.9e-25	86.7	2	279-422
756	ig	Immunoglobulin domain	1.3e-06	27.6	1	43-102
756	ig	Immunoglobulin domain	2.2e-05	23.0	2	137-198
756	FYRN	F/Y-rich N-terminus	0.55	5.3	1	181-200
756	ig	Immunoglobulin domain	6.5e-09	36.2	3	242-299
756	ig	Immunoglobulin domain	2.3e-05	22.9	4	339-388
756	ig	Immunoglobulin domain	2.9e-08	33.8	5	424-481
756	ig	Immunoglobulin domain	7.7e-07	28.5	6	514-579
756	fn3	Fibronectin type III domain	7.7e-23	81.1	1	598-687
756	fn3	Fibronectin type III domain	9.1e-08	28.7	2	700-790
756	fn3	Fibronectin type III domain	9.3e-17	60.0	3	802-891
756	fn3	Fibronectin type III domain	1.6e-09	34.8	4	903-986
757	LRR	Leucine Rich Repeat	0.29	7.0	1	52-75
757	LRR	Leucine Rich Repeat	0.003	13.7	2	76-99
757	LRR	Leucine Rich Repeat	4e-05	20.0	3	100-123
757	LRR	Leucine Rich Repeat	0.021	10.8	4	124-147
757	LRR	Leucine Rich Repeat	3e-05	20.4	5	148-171
757	LRR	Leucine Rich Repeat	0.00019	17.8	6	172-195
757	FliD	Flagellar hook-associated protein 2	0.96	1.2	1	194-209
757	LRR	Leucine Rich Repeat	0.16	7.8	7	196-216
757	LRRCT	Leucine rich repeat C-terminal domain	9.3e-10	31.0	1	240-285
757	ig	Immunoglobulin domain	9.4e-09	35.6	1	301-359
757	fn3	Fibronectin type III domain	0.00045	15.9	1	444-496
758	LRR	Leucine Rich Repeat	0.29	7.0	1	52-75
758	LRR	Leucine Rich Repeat	0.003	13.7	2	76-99
758	LRR	Leucine Rich Repeat	4e-05	20.0	3	100-123
758	LRR	Leucine Rich Repeat	0.021	10.8	4	124-147
758	LRR	Leucine Rich Repeat	3e-05	20.4	5	148-171
758	LRR	Leucine Rich Repeat	0.00019	17.8	6	172-195
758	FliD	Flagellar hook-associated protein 2	0.96	1.2	1	194-209
758	LRR	Leucine Rich Repeat	0.16	7.8	7	196-216
758	LRRCT	Leucine rich repeat C-terminal domain	9.3e-10	31.0	1	240-285
758	ig	Immunoglobulin domain	9.4e-09	35.6	1	301-359
758	fn3	Fibronectin type III domain	0.013	10.8	1	466-500

505
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
759	Serendipity_A	Serendipity locus alpha protein (SRY-	0.35	2.3	1	98-106
759	EGF	EGF-like domain	0.76	5.8	1	111-133
759	SEA	SEA domain	4.9e-06	22.1	1	168-237
759	ig	Immunoglobulin domain	9.8e-07	28.1	1	286-352
759	AIG2	AIG2-like family	0.81	2.4	1	329-340
759	ig	Immunoglobulin domain	0.33	7.4	2	485-547
759	60KD_IMP	60Kd inner membrane protein	0.64	3.1	1	502-523
759	Atracotoxin	Delta Atracotoxin	0.31	6.4	1	628-642
759	CAS_CSE1	CAS/CSE protein, C-terminus	0.28	5.8	1	902-915
759	GPS	Latrophilin/CL-1-like GPS domain	2e-14	54.5	1	950-1002
759	7tm_2	7 transmembrane receptor (Secretin fa	6.4e-21	73.4	1	1009-1273
759	ATP-synt_G	Mitochondrial ATP synthase g subunit	0.66	3.9	1	1267-1279
759	SH	Viral small hydrophobic protein	0.63	4.1	1	1273-1292
760	TFIIS	Transcription factor S-II (TFIIS)	0.27	6.6	1	107-117
760	zf-C2H2	1/13 93 101 .. 16 24	0.00013	23.2	2	107-129
760	zf-C2H2	1/13 93 101 .. 16 24	3.2e-06	29.6	3	135-157
760	Ribosomal_L19e	Ribosomal protein L19e	0.59	3.9	1	143-161
760	TFIIS	Transcription factor S-II (TFIIS)	0.08	8.4	3	163-173
760	zf-C2H2	1/13 93 101 .. 16 24	1.5e-05	26.9	4	163-185
760	XPA_N	XPA protein N-terminal	0.37	6.2	3	188-200
760	TFIIS	Transcription factor S-II (TFIIS)	0.08	8.4	4	191-201
760	zf-C2H2	1/13 93 101 .. 16 24	4.3e-06	29.1	5	191-213
760	XPA_N	XPA protein N-terminal	0.15	7.6	4	216-229
760	TFIIS	Transcription factor S-II (TFIIS)	0.31	6.4	5	219-229
760	zf-C2H2	1/13 93 101 .. 16 24	2.4e-06	30.1	6	219-241
760	XPA_N	XPA protein N-terminal	0.45	5.9	5	244-256
760	TFIIS	Transcription factor S-II (TFIIS)	0.013	11.2	6	247-257
760	zf-C2H2	1/13 93 101 .. 16 24	9.1e-07	31.8	7	247-269
760	XPA_N	XPA protein N-terminal	0.29	6.6	6	272-284
760	zf-C2H2	1/13 93 101 .. 16 24	3.8e-08	37.2	8	275-297
760	zf-BED	BED zinc finger	0.13	7.1	3	276-298
760	zf-C2H2	1/13 93 101 .. 16 24	3e-06	29.7	9	303-325
760	TFIIS	Transcription factor S-II (TFIIS)	0.019	10.6	9	331-341
760	zf-C2H2	1/13 93 101 .. 16 24	3.1e-06	29.7	10	331-353
760	zf-C2H2	1/13 93 101 .. 16 24	2.7e-07	33.9	11	359-381
760	zf-BED	BED zinc finger	0.63	4.8	4	360-382
760	PqiA	Paraquat-inducible protein A	0.55	4.0	2	378-409
760	XPA_N	9/11 356 366 .. 1 11	0.22	7.0	10	384-396
760	zf-C2H2	1/13 93 101 .. 16 24	8.8e-08	35.9	12	387-409
760	TFIIS	Transcription factor S-II (TFIIS)	0.036	9.7	12	415-425
760	zf-C2H2	1/13 93 101 .. 16 24	0.028	13.7	13	415-437
761	C1q	C1q domain	0.77	4.7	1	104-116
761	DUF127	Protein of unknown function DUF127	0.81	2.3	1	134-143
761	Hydrolase	haloacid dehalogenase-like hydrolase	0.53	4.3	1	176-189
761	Hydrolase	haloacid dehalogenase-like hydrolase	0.27	5.3	2	443-477
761	Hydrolase	haloacid dehalogenase-like hydrolase	0.65	4.0	3	543-620
761	PgpA	Phosphatidylglycerophosphatase A	0.96	2.9	1	745-760
761	DUF418	Protein of unknown function (DUF418)	0.15	6.0	1	833-887
763	zf-HIT	HIT zinc finger	0.21	8.5	1	161-179
763	zf-C2H2	Zinc finger, C2H2 type	0.0099	15.5	1	170-193

506
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
764	FHA	FHA domain	0.024	11.6	1	25-90
764	HIT	HIT domain	6e-05	15.8	1	181-235
764	DcpS	Scavenger mRNA decapping enzyme (DcpS)	0.0099	8.2	1	210-271
764	DUF369	Domain of unknown function (DUF369)	0.35	4.8	1	219-239
764	zf-C2H2	Zinc finger, C2H2 type	0.026	13.9	1	317-339
767	Cwf_Cwc_15	Cwf15/Cwc15 cell cycle control protei	8.6e-161	544.3	1	1-229
767	DUF692	Protein of unknown function (DUF692)	0.91	2.6	1	127-148
768	SRCR	Scavenger receptor cysteine-rich doma	9.2e-36	127.9	1	32-129
768	SRCR	Scavenger receptor cysteine-rich doma	6.5e-15	54.2	2	142-247
768	Lysyl oxidase	Lysyl oxidase	1.2e-80	278.1	1	251-359
769	RHS	RHS protein	0.83	4.8	1	31-43
769	GatB	PET112 family, C terminal region	0.41	5.8	1	64-86
769	Glyco_transf_8	Glycosyl transferase family 8	1.9e-10	40.1	1	65-227
769	Phage_holin_4	Holin family	0.84	4.0	1	269-282
770	WD40	WD domain, G-beta repeat	0.5	6.4	2	169-194
770	WD40	WD domain, G-beta repeat	5.2e-06	23.8	3	225-251
770	DUF130	Domain of unknown function DUF130	0.074	5.9	1	241-255
770	WD40	WD domain, G-beta repeat	0.35	7.0	4	374-401
771	TPR	TPR Domain	0.27	7.6	1	190-214
773	CTP_transf_1	Cytidylyltransferase family	3.3e-125	426.1	1	69-400
773	DAG_PE-bind	Phorbol esters/diacylglycerol binding	0.28	7.3	1	166-180
773	Pyridox_oxidase	Pyridoxamine 5'-phosphate oxidase	0.34	2.7	1	326-334
773	KIX	KIX domain	0.48	5.9	1	415-435
774	CBM_20	Starch binding domain	0.078	8.5	1	86-105
774	WD40	WD domain, G-beta repeat	3.9e-08	31.2	1	165-203
775	TACC	Transforming acidic coiled-coil-conta	0.43	3.9	1	312-334
775	bZIP	1/2 308 325 .. 48 65	0.39	5.9	2	408-438
776	Tweety	Tweety	3.4e-74	256.6	1	21-413
779	HesB-like	HesB-like domain	2.8e-41	132.5	1	49-151
780	ig	Immunoglobulin domain	0.015	12.4	1	2-57
780	ig	Immunoglobulin domain	0.00033	18.6	2	96-155
781	Mpv17_PMP22	Mpv17 / PMP22 family	8e-14	51.5	1	129-191
781	Adenovirus_PX	Adenovirus late L2 mu core protein (P	0.65	5.4	1	133-152
782	sic	sic protein	0.1	3.9	1	184-239
783	Collagen	Collagen triple helix repeat (20 copi	5.5e-07	28.5	1	13-51
783	Collagen	Collagen triple helix repeat (20 copi	0.044	10.1	2	59-81
783	Collagen	Collagen triple helix repeat (20 copi	0.014	12.0	3	86-104
783	Collagen	Collagen triple helix repeat (20 copi	0.029	10.7	4	106-127
783	Collagen	Collagen triple helix repeat (20 copi	0.013	12.1	5	132-150
783	Collagen	Collagen triple helix repeat (20 copi	0.04	10.3	6	152-173
783	Collagen	Collagen triple helix repeat (20 copi	0.013	12.1	7	175-196
783	Collagen	Collagen triple helix repeat (20 copi	2.5e-07	29.8	8	198-237
783	S-AdoMet_syntD3	S-adenosylmethionine synthetase, C-te	0.29	4.1	1	232-247
783	vwa	von Willebrand factor type A domain	1.2e-46	149.1	1	266-448
783	Kunitz_BPTI	Kunitz/Bovine pancreatic trypsin inhi	2.2e-23	71.1	1	540-590
784	DUF388	Domain unknown function (DUF388)	0.047	8.8	1	1-18
784	Mtap_PNP	Phosphorylase family 2	0.26	5.1	1	1-18
784	Sterol_desat	Sterol desaturase	1.8e-48	164.1	1	57-263
785	ig	Immunoglobulin domain	0.0011	16.7	1	116-176

507

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
785	malic	Malic enzyme, N-terminal domain	0.76	2.6	1	203-209
785	ig	Immunoglobulin domain	3.6e-10	40.9	2	331-391
785	APOBEC_C	APOBEC-like C-terminal domain	0.88	5.2	1	446-474
785	enolase	Enolase, C-terminal TIM barrel domain	0.26	4.4	1	600-618
785	RNA_pol_Rpb1_5	RNA polymerase Rpb1, domain 5	0.65	1.5	1	1034-1489
785	ig	Immunoglobulin domain	4.9e-05	21.7	3	1363-1415
785	sigma70_r3	Sigma-70 region 3	0.6	5.7	1	1461-1481
785	ig	Immunoglobulin domain	5.2e-08	32.9	4	1552-1613
786	RNA_helicase	RNA helicase	0.00029	15.0	1	30-53
786	AAA	ATPase family associated with various	0.00038	13.8	1	32-48
786	NACHT	NACHT domain	0.0022	12.0	1	34-56
786	ATP-bind	Conserved hypothetical ATP binding pr	0.64	3.5	1	35-46
786	NB-ARC	NB-ARC domain	0.62	2.7	1	35-50
786	ADK	Adenylate kinase	2.2e-05	19.0	1	67-114
786	ADK	Adenylate kinase	0.12	6.2	2	127-160
786	ZZ	Zinc finger, ZZ type	0.097	8.8	1	146-157
786	SRP54	SRP54-type protein, GTPase domain	0.3	5.9	1	390-408
786	SKI	Shikimate kinase	0.12	6.4	1	392-413
786	ATP-bind	Conserved hypothetical ATP binding pr	0.83	3.1	2	396-413
786	RHD3	Root hair defective 3 GTP-binding pro	0.039	5.3	1	397-411
786	CoaE	Dephospho-CoA kinase	0.12	6.4	1	402-421
786	Thymidylate_kin	Thymidylate kinase	0.81	2.1	1	402-418
788	SH3	SH3 domain	2.3e-14	55.4	1	1-56
789	SH3	SH3 domain	1.5e-15	59.8	1	73-129
790	TIMP	Tissue inhibitor of metalloproteinase	1.2e-89	243.9	1	20-116
790	Phyto_Pns9_10	Phytoreovirus nonstructural protein P	0.44	3.0	1	102-108
791	DUF716	Family of unknown function (DUF716)	0.93	3.4	1	26-54
791	DcuC	C4-dicarboxylate anaerobic carrier	0.4	4.3	1	27-48
791	FLO_LFY	Floricaula / Leafy protein	0.22	2.7	1	127-140
791	lectin_c	Lectin C-type domain	1.9e-07	31.5	1	162-267
792	UDPGT	UDP-glucuronosyl and UDP-glucosyl tra	7.9e-258	866.7	1	24-447
792	Pox_E8	Poxvirus E8 protein	0.81	3.1	1	56-70
792	Glyco_tran_28_C	Glycosyltransferase family 28 C-termi	0.06	7.5	1	292-314
793	TRAPP_Bet3	Transport protein particle (TRAPP) co	1.1e-67	235.0	1	6-173
794	PCMT	Protein-L-isoaspartate(D-aspartate) O	0.0055	11.0	1	74-113
794	Ubie_methyltran	ubiE/COQ5 methyltransferase family	1.9e-05	18.5	1	161-182
794	Methyltransf_8	Hypothetical methyltransferase	0.04	7.5	1	168-182
795	Brix	Brix domain	2e-88	303.9	1	1-248
795	PDZ	PDZ domain (Also known as DHR or GLGF	0.3	6.3	1	246-273
795	7tm_1	7 transmembrane receptor (rhodopsin f	3e-42	125.3	1	444-671
795	Lip_A_acyltrans	Bacterial lipid A biosynthesis acyltr	0.4	4.3	1	532-558
795	ACPS	4'-phosphopantetheinyl transferase su	0.72	3.3	1	585-600
796	ig	Immunoglobulin domain	0.0042	14.5	1	33-110
797	ig	Immunoglobulin domain	0.0042	14.5	1	33-110
798	ig	Immunoglobulin domain	0.0042	14.5	1	33-110
798	FliL	Flagellar basal body-associated prote	0.029	9.2	1	170-203
798	DcuC	C4-dicarboxylate anaerobic carrier	0.044	7.9	1	174-193
799	PH	PH domain	1.9e-21	72.0	1	14-112

508

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
800	Ifi-6-16	Interferon-induced 6-16 family	1.1e-41	144.4	1	10-87
800	GLTT	GLTT repeat (6 copies)	0.18	7.7	1	14-42
800	CRCB	CrcB-like protein	0.18	7.1	1	70-88
801	Ifi-6-16	Interferon-induced 6-16 family	3.7e-46	159.7	1	17-99
801	GLTT	GLTT repeat (6 copies)	0.18	7.7	1	26-54
801	CRCB	CrcB-like protein	0.18	7.1	1	82-100
802	ank	Ankyrin repeat	1	5.7	1	338-367
802	RmuC	RmuC family	0.49	3.9	1	621-657
804	ig	Immunoglobulin domain	0.0002	19.4	1	35-111
804	DUF708	Protein of unknown function (DUF708)	0.27	5.6	1	230-246
804	CDC50	LEM3 (ligand-effect modulator 3) fami	0.049	6.6	1	231-258
806	EGF	EGF-like domain	0.0019	15.2	1	60-95
807	EGF	EGF-like domain	0.0019	15.2	1	60-95
808	EGF	EGF-like domain	0.0019	15.2	1	60-95
809	PI3_PI4_kinase	Phosphatidylinositol 3- and 4-kinase	0.89	3.6	1	6-35
809	ig	Immunoglobulin domain	4.9e-06	25.4	1	109-171
811	Alpha_adaptin_C	Alpha adaptin AP2, C-terminal domain	0.061	5.2	1	92-104
811	MHC_I	Class I Histocompatibility antigen, d	0.021	9.1	2	120-205
812	ig	Immunoglobulin domain	3.7e-10	40.9	2	78-137
812	ig	Immunoglobulin domain	0.0018	15.9	3	176-237
812	ig	Immunoglobulin domain	3.7e-08	33.4	4	274-335
812	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	291-347
812	OapA	Opacity-associated protein A	0.44	2.4	1	300-322
812	ig	Immunoglobulin domain	0.0012	16.6	5	369-430
812	ig	Immunoglobulin domain	7.7e-07	28.5	6	465-529
813	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.65	2.6	1	55-63
813	LRRCT	Leucine rich repeat C-terminal domain	0.15	5.9	1	61-85
813	DUF909	Bacterial protein of unknown function	0.4	5.7	1	237-256
813	ig	Immunoglobulin domain	0.0047	14.3	1	295-358
813	ig	Immunoglobulin domain	1.2e-08	35.2	2	393-452
813	Nol1_Nop2_Sun	NOL1/NOP2/sun family	0.28	4.1	1	629-671
813	ig	Immunoglobulin domain	1.2e-05	24.0	3	1468-1530
813	ig	Immunoglobulin domain	1.1e-06	27.9	4	1565-1627
813	ig	Immunoglobulin domain	6.2e-09	36.3	5	1662-1724
813	CD2	T-cell surface antigen CD2 protein	0.19	3.9	1	1701-1749
813	ig	Immunoglobulin domain	2.6e-09	37.7	6	1761-1823
813	ig	Immunoglobulin domain	8.7e-06	24.5	7	1858-1926
813	ig	Immunoglobulin domain	3.7e-10	40.9	8	1961-2020
813	ig	Immunoglobulin domain	0.0018	15.9	9	2059-2120
813	ig	Immunoglobulin domain	3.7e-08	33.4	10	2157-2218
813	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	2174-2230
813	OapA	Opacity-associated protein A	0.44	2.4	1	2183-2205

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
813	ig	Immunoglobulin domain	0.0012	16.6	11	2252-2313
813	ig	Immunoglobulin domain	7.7e-07	28.5	12	2348-2412
814	DUF126	Protein of unknown function DUF126	0.08	6.1	1	1-9
814	LRRNT	Leucine rich repeat N-terminal domain	3e-07	26.4	1	28-56
814	LRR	Leucine Rich Repeat	0.0074	12.4	1	58-81
814	Phage_holin_4	Holin family	0.73	4.2	1	69-88
814	LRR	Leucine Rich Repeat	0.00054	16.2	2	82-105
814	LRR	Leucine Rich Repeat	0.005	12.9	3	106-129
814	LRR	Leucine Rich Repeat	0.00025	17.3	4	130-153
814	LRR	Leucine Rich Repeat	0.00088	15.5	5	154-177
814	LRR	Leucine Rich Repeat	0.0028	13.8	6	186-209
814	LRRCT	Leucine rich repeat C-terminal domain	2.4e-13	42.0	1	219-280
814	DUF909	Bacterial protein of unknown function	0.4	5.7	1	432-451
814	ig	Immunoglobulin domain	0.0047	14.3	1	490-553
814	ig	Immunoglobulin domain	1.2e-08	35.2	2	588-647
814	Nol1_Nop2_Sun	NOL1/NOP2/sun family	0.28	4.1	1	824-866
814	ig	Immunoglobulin domain	1.2e-05	24.0	3	1663-1725
814	ig	Immunoglobulin domain	1.1e-06	27.9	4	1760-1822
814	ig	Immunoglobulin domain	6.2e-09	36.3	5	1857-1919
814	CD2	T-cell surface antigen CD2 protein	0.19	3.9	1	1896-1944
814	ig	Immunoglobulin domain	2.6e-09	37.7	6	1956-2018
814	ig	Immunoglobulin domain	8.7e-06	24.5	7	2053-2121
814	ig	Immunoglobulin domain	3.7e-10	40.9	8	2156-2215
814	ig	Immunoglobulin domain	0.0018	15.9	9	2254-2315
814	ig	Immunoglobulin domain	3.7e-08	33.4	10	2352-2413
814	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	2369-2425
814	OapA	Opacity-associated protein A	0.44	2.4	1	2378-2400
814	ig	Immunoglobulin domain	0.0012	16.6	11	2447-2508
814	ig	Immunoglobulin domain	7.7e-07	28.5	12	2543-2607
816	Apolipoprotein	Apolipoprotein A1/A4/E family	2.3e-11	42.3	1	93-168
816	DUF260	Protein of unknown function DUF260	0.64	3.5	1	94-107
816	Adeno_PIX	Adenovirus hexon-associated protein (0.49	4.4	1	95-110
816	BcrAD_BadFG	BadF/BadG/BcrA/BcrD ATPase family	0.12	6.2	1	134-180
816	Apolipoprotein	Apolipoprotein A1/A4/E family	0.011	10.5	2	172-258
816	MM_CoA_mutase	Methylmalonyl-CoA mutase	0.84	1.9	1	264-306
817	Apolipoprotein	Apolipoprotein A1/A4/E family	2.3e-11	42.3	1	93-168
817	DUF260	Protein of unknown function DUF260	0.64	3.5	1	94-107
817	Adeno_PIX	Adenovirus hexon-associated protein (0.49	4.4	1	95-110

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
817	BcrAD_BadFG	BadF/BadG/BcrA/BcrD ATPase family	0.12	6.2	1	134-180
817	Apolipoprotein	Apolipoprotein A1/A4/E family	0.011	10.5	2	172-258
817	MM_CoA_mutase	Methylmalonyl-CoA mutase	0.84	1.9	1	264-306
818	DUF717	Protein of unknown function (DUF717)	1	4.0	1	109-121
818	MHC_I	Class I Histocompatibility antigen, d	0.69	3.7	1	226-239
819	Pox_D5	Poxvirus D5 protein-like	1	2.2	1	16-28
819	phoslip	Phospholipase A2	3.4e-49	172.4	1	21-145
819	RFX_DNA_binding	RFX DNA-binding domain	0.84	2.9	1	50-57
821	MR_MLE_N	Mandelate racemase / muconate lactoni	1.6e-05	17.0	1	9-112
821	Peptidase_S26	Signal peptidase I	0.38	3.8	1	54-84
821	CheR_N	CheR methyltransferase, all-alpha dom	0.4	6.7	1	58-74
821	MR_MLE	Mandelate racemase / muconate lactoni	2.5e-08	29.9	1	191-253
822	NAP	Nucleosome assembly protein (NAP)	6e-191	644.5	1	12-285
822	GAT	GAT domain	0.27	4.9	1	114-126
822	DUF115	Protein of unknown function DUF115	0.76	3.8	1	116-143
823	PP2C	Protein phosphatase 2C	3.4e-72	250.0	1	107-383
824	vwc	von Willebrand factor type C domain	2.2e-10	37.8	1	103-157
824	vwc	von Willebrand factor type C domain	4.7e-09	33.1	2	160-205
824	TILa	TILa domain	0.24	6.3	2	183-200
825	7tm_1	7 transmembrane receptor (rhodopsin f	1.4e-28	84.3	1	1-173
826	7tm_1	7 transmembrane receptor (rhodopsin f	4.5e-49	145.7	1	40-287
827	EGF	EGF-like domain	0.0067	13.2	1	35-62
827	DSL	Delta serrate ligand	0.48	4.7	1	47-62
828	Pox_A46	Poxvirus A46 family	0.55	2.5	1	1-15
828	ExoD	Exopolysaccharide synthesis, ExoD	0.82	2.4	1	64-87
828	RhoGAP	RhoGAP domain	1.3e-53	182.4	1	101-250
828	Sec6	Exocyst complex component Sec6	0.97	1.8	1	184-207
829	CUB	CUB domain	1.1e-33	112.6	1	5-102
830	CUB	CUB domain	1.1e-33	112.6	1	5-102
831	myosin_head	Myosin head (motor domain)	8.8e-76	257.2	1	37-299
831	ATP_bind2	P-loop ATPase protein family	0.16	4.9	1	126-139
831	PRK	Phosphoribulokinase / Uridine kinase	0.14	5.2	1	128-139
832	myosin_head	Myosin head (motor domain)	4.1e-90	306.1	1	37-387
832	ATP_bind2	P-loop ATPase protein family	0.16	4.9	1	126-139
832	PRK	Phosphoribulokinase / Uridine kinase	0.14	5.2	1	128-139
834	7tm_5	7TM chemoreceptor	0.17	1.1	1	37-49
834	kazal	Kazal-type serine protease inhibitor	8.4e-08	33.5	1	139-183
834	thyroglobulin_1	Thyroglobulin type-1 repeat	4.1e-21	80.3	1	316-379
835	Micro_A_star	Microvirus A* protein	0.16	5.3	1	410-426
835	Coronavirus_5	Coronavirus gene 5 protein	0.91	3.0	1	540-553
835	RPEL	RPEL repeat	0.81	5.4	1	540-550
836	Micro_A_star	Microvirus A* protein	0.16	5.3	1	410-426
836	Coronavirus_5	Coronavirus gene 5 protein	0.91	3.0	1	540-553
836	RPEL	RPEL repeat	0.81	5.4	1	540-550
837	BEX	Brain expressed X-linked like family	9.8e-86	266.4	1	14-125
837	ChaC	ChaC-like protein	0.2	4.5	1	67-92
837	IlvC	Acetohydroxy acid isomeroreductase, c	0.14	5.9	1	68-97
838	LRRNT	Leucine rich repeat N-terminal domain	4.1e-05	19.3	1	31-59
838	LRR	Leucine Rich Repeat	0.045	9.7	1	61-84
838	LRR	Leucine Rich Repeat	0.0026	13.9	3	109-132
838	LRR	Leucine Rich Repeat	0.002	14.3	4	133-156
838	LRR	Leucine Rich Repeat	0.0034	13.5	5	157-180

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
838	LRR	Leucine Rich Repeat	0.00019	17.8	6	181-204
838	LRR	Leucine Rich Repeat	6.4e-05	19.3	7	205-228
838	LRR	Leucine Rich Repeat	3.4e-05	20.2	8	229-252
838	LRR	Leucine Rich Repeat	0.59	6.0	9	253-276
838	LRR	Leucine Rich Repeat	9.3e-05	18.8	10	277-300
838	LRR	Leucine Rich Repeat	0.0022	14.1	11	301-324
838	Scramblase	Scramblase	0.76	1.7	1	313-322
838	LRR	Leucine Rich Repeat	0.0001	18.6	12	326-349
838	LRRCT	Leucine rich repeat C-terminal domain	4.3e-13	41.2	1	359-405
838	UPF0118	Domain of unknown function DUF20	1	2.9	1	533-556
840	dUTPase	dUTPase	0.34	6.2	1	343-362
841	ank	Ankyrin repeat	0.00082	16.7	1	1-27
841	MM_CoA_mutase	Methylmalonyl-CoA mutase	0.85	1.9	1	9-43
841	ank	Ankyrin repeat	7.1e-07	27.7	2	29-61
841	ank	Ankyrin repeat	2.3e-09	36.6	3	130-162
841	ank	Ankyrin repeat	2.2e-10	40.3	4	164-196
841	Myc_N_term	Myc amino-terminal region	0.27	3.6	1	514-541
841	SAM	SAM domain (Sterile alpha motif)	1.3e-06	25.0	1	588-640
842	DUF370	Domain of unknown function (DUF370)	1	3.5	1	21-36
842	ApoL	Apolipoprotein L	3.1e-195	658.7	1	43-345
842	HupH_C	HupH hydrogenase expression protein,	0.99	2.7	1	116-131
842	DUF710	Family of unknown function (DUF710)	0.48	5.0	1	297-337
843	DUF370	Domain of unknown function (DUF370)	1	3.5	1	21-36
843	ApoL	Apolipoprotein L	1.7e-194	656.3	1	43-345
843	HupH_C	HupH hydrogenase expression protein,	0.99	2.7	1	116-131
843	DUF710	Family of unknown function (DUF710)	0.48	5.0	1	297-337
844	Uteroglobin	Uteroglobin family	1	3.3	1	1-16
844	DUF84	Protein of unknown function DUF84	0.098	5.9	1	8-22
844	DUF960	Staphylococcal protein of unknown fun	0.78	3.7	1	38-63
844	Tail_X	Phage Tail Protein X	0.35	5.8	1	45-56
844	LysM	LysM domain	0.36	6.9	1	48-56
844	ig	Immunoglobulin domain	3e-07	30.0	1	53-110
844	ig	Immunoglobulin domain	1.8e-07	30.9	2	150-216
844	ig	Immunoglobulin domain	2.9e-08	33.8	3	255-310
844	ig	Immunoglobulin domain	4.6e-07	29.3	4	350-417
845	Uteroglobin	Uteroglobin family	1	3.3	1	1-16
845	DUF84	Protein of unknown function DUF84	0.098	5.9	1	8-22
845	DUF960	Staphylococcal protein of unknown fun	0.78	3.7	1	38-63
845	Tail_X	Phage Tail Protein X	0.35	5.8	1	45-56
845	LysM	LysM domain	0.36	6.9	1	48-56
845	ig	Immunoglobulin domain	3e-07	30.0	1	53-110
845	ig	Immunoglobulin domain	1.8e-07	30.9	2	150-216
845	ig	Immunoglobulin domain	2.9e-08	33.8	3	255-310
845	ig	Immunoglobulin domain	4.6e-07	29.3	4	350-417
845	ig	Immunoglobulin domain	1.1e-07	31.6	5	456-516
845	ig	Immunoglobulin domain	8.8e-05	20.8	6	553-617
845	APS_kinase	Adenylylsulphate kinase	0.67	2.8	1	593-609
845	fn3	Fibronectin type III domain	0.75	4.7	1	656-733
845	MAM	MAM domain	6.7e-77	265.6	1	753-918

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
845	E2F_TDP	Transcription factor E2F/dimerisation	0.56	3.7	1	761-787
846	zf-PARP	Poly(ADP-ribose) polymerase and DNA-L	0.61	5.0	1	38-54
846	Albicidin_res	Albicidin resistance domain	0.49	6.1	1	290-297
846	SPDY	Domain of unknown function (DUF317)	0.37	5.2	1	361-374
846	CBF	CBF/Mak21 family	0.00014	14.4	1	417-450
847	CNH	CNH domain	0.00087	13.7	1	164-217
847	NHL	NHL repeat	0.14	9.4	1	204-229
847	Coprogen_oxidas	Coproporphyrinogen III oxidase	0.26	1.9	1	231-246
847	Clathrin	Region in Clathrin and VPS	0.0094	11.5	1	404-445
847	ENTH	ENTH domain	0.31	5.7	1	794-807
847	C2	C2 domain	2.2e-18	63.6	1	797-876
847	PLA2_B	Lysophospholipase catalytic domain	9.1e-51	178.0	1	1108-1317
847	DUF188	Uncharacterized BCR, Yail/YqxD family	0.9	2.9	1	1314-1325
847	TAP42	TAP42-like family	1	2.0	1	1408-1413
847	PLA2_B	Lysophospholipase catalytic domain	1.2e-12	43.6	2	1429-1551
848	ENTH	ENTH domain	0.31	5.7	1	43-56
848	C2	C2 domain	2.2e-18	63.6	1	46-125
848	PLA2_B	Lysophospholipase catalytic domain	2.4e-53	187.1	1	357-566
848	DUF188	Uncharacterized BCR, Yail/YqxD family	0.9	2.9	1	563-574
848	TAP42	TAP42-like family	1	2.0	1	657-662
848	PLA2_B	Lysophospholipase catalytic domain	1.2e-12	43.6	2	678-800
849	SNF7	SNF7	1.3e-54	191.6	1	18-178
849	GatB_N	PET112 family, N terminal region	0.2	4.6	1	135-146
849	Interleukin_13	Interleukin-13	0.24	6.5	1	156-167
850	p450	Cytochrome P450	2.9e-05	15.6	1	25-112
850	Phage_attach	Phage Head-Tail Attachment	0.97	1.6	1	69-80
851	ig	Immunoglobulin domain	8e-09	35.9	1	48-105
851	ig	Immunoglobulin domain	1.5e-12	49.8	2	169-227
851	ig	Immunoglobulin domain	2.3e-06	26.7	3	265-344
851	CD36	CD36 family	0.38	3.9	1	377-402
851	Neur_chan_mem b	Neurotransmitter-gated ion-channel tr	0.69	2.3	1	392-401
852	ig	Immunoglobulin domain	8e-09	35.9	1	44-101
852	ig	Immunoglobulin domain	1.5e-12	49.8	2	165-223
852	ig	Immunoglobulin domain	2.3e-06	26.7	3	261-340
852	CD36	CD36 family	0.38	3.9	1	373-398
852	Neur_chan_mem b	Neurotransmitter-gated ion-channel tr	0.69	2.3	1	388-397
853	ig	Immunoglobulin domain	8e-09	35.9	1	44-101
853	bZIP_Maf	bZIP Maf transcription factor	0.4	4.3	1	101-127
854	C2	C2 domain	1.8e-39	134.8	1	158-245
854	C2	C2 domain	8.3e-37	125.8	2	289-377
855	DUF1058	Protein of unknown function (DUF1058)	0.49	2.3	1	79-92
855	Pep_M12B_prop ep	Reprolysin family propeptide	7.2e-06	18.8	1	154-222
855	Reprolysin	Reprolysin (M12B) family zinc metallo	9.5e-18	66.0	2	313-456

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
855	Mu-conotoxin	Mu-Conotoxin	0.94	4.6	1	356-377
855	Astacin	Astacin (Peptidase family M12A)	0.65	3.4	1	389-402
855	fn2	Fibronectin type II domain	0.59	3.4	1	445-451
855	tsp_1	Thrombospondin type 1 domain	3e-16	55.9	1	546-596
855	ADAM_spacer1	ADAM-TS Spacer 1	1.6e-49	174.7	1	702-813
855	DSL	Delta serrate ligand	0.38	5.0	1	794-812
855	tsp_1	Thrombospondin type 1 domain	0.0007	14.6	2	832-844
855	zf-NF-X1	NF-X1 type zinc finger	0.0071	8.8	2	873-895
855	tsp_1	Thrombospondin type 1 domain	0.0028	12.7	3	888-909
855	tsp_1	Thrombospondin type 1 domain	8.2e-08	27.8	4	945-995
855	Reo_sigmaC	Reovirus sigma C capsid protein	0.73	2.0	1	1216-1224
855	UPF0051	Uncharacterized protein family (UPF00	0.0073	8.9	1	1284-1297
855	tsp_1	Thrombospondin type 1 domain	0.01	10.8	5	1321-1364
855	tsp_1	Thrombospondin type 1 domain	0.0037	12.3	7	1429-1471
855	tsp_1	Thrombospondin type 1 domain	3.4e-05	19.0	8	1474-1530
856	Ifi-6-16	Interferon-induced 6-16 family	3.5e-07	26.2	1	21-44
856	CRCB	CrcB-like protein	0.18	7.1	1	27-45
857	GHMP_kinases	GHMP kinases putative ATP-binding pro	0.55	1.9	1	81-129
857	abhydrolase	alpha/beta hydrolase fold	0.02	9.2	1	161-214
857	lipase	Lipase	0.64	3.7	1	185-213
857	abhydrolase	alpha/beta hydrolase fold	0.0083	10.5	2	254-324
857	DLH	Dienelactone hydrolase family	0.4	3.6	1	256-283
857	LIP	Secretory lipase	0.012	8.6	1	265-290
857	UPF0227	Uncharacterised protein family (UPF02	0.38	4.9	1	266-296
857	abhydrolase_2	Phospholipase/Carboxylesterase	0.015	10.1	1	267-290
857	Peptidase_M10_N	Matrix metalloprotease, N-terminal do	0.63	2.5	1	296-317
858	GHMP_kinases	GHMP kinases putative ATP-binding pro	0.55	1.9	1	74-122
858	abhydrolase	alpha/beta hydrolase fold	0.02	9.2	1	154-207
858	lipase	Lipase	0.64	3.7	1	178-206
858	abhydrolase	alpha/beta hydrolase fold	0.0083	10.5	2	247-317
858	DLH	Dienelactone hydrolase family	0.4	3.6	1	249-276
858	LIP	Secretory lipase	0.012	8.6	1	258-283
858	UPF0227	Uncharacterised protein family (UPF02	0.38	4.9	1	259-289
858	abhydrolase_2	Phospholipase/Carboxylesterase	0.015	10.1	1	260-283
858	Peptidase_M10_N	Matrix metalloprotease, N-terminal do	0.63	2.5	1	289-310
859	H-kinase_dim	Signal transducing histidine kinase,	0.25	6.8	1	15-55
859	Collagen	Collagen triple helix repeat (20 copi	4.8e-08	32.5	1	244-284
859	Collagen	Collagen triple helix repeat (20 copi	3.3e-05	21.8	2	285-320
859	SRCR	Scavenger receptor cysteine-rich doma	6.6e-22	78.9	1	336-433
859	MBD	Methyl-CpG binding domain	0.52	4.9	1	365-389
860	CobS	Cobalamin-5-phosphate synthase	0.43	3.4	1	45-58
860	LGT	Prolipoprotein diacylglycerol transfe	0.084	6.6	1	64-85
860	Collagen	Collagen triple helix repeat (20 copi	2.6e-07	29.7	1	304-344
860	Collagen	Collagen triple helix repeat (20 copi	3.3e-05	21.8	2	345-380
860	SRCR	Scavenger receptor cysteine-rich doma	2.7e-34	122.7	1	396-493

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
862	TFIIS	Transcription factor S-II (TFIIS)	0.73	5.1	1	192-202
862	zf-C2H2	Zinc finger, C2H2 type	3.5e-05	25.4	1	192-214
862	zf-C2H2	Zinc finger, C2H2 type	1.3e-06	31.2	2	220-242
862	zf-BED	BED zinc finger	0.33	5.7	1	222-243
862	mRNA_cap_enzyme	mRNA capping enzyme, catalytic domain	0.56	0.5	1	245-260
862	XPA_N	XPA protein N-terminal	0.78	5.1	2	245-257
862	zf-C2H2	Zinc finger, C2H2 type	2.9e-07	33.8	3	248-270
862	TFIIS	Transcription factor S-II (TFIIS)	0.89	4.8	3	276-286
862	zf-C2H2	Zinc finger, C2H2 type	2e-06	30.4	4	276-298
862	zf-C2H2	Zinc finger, C2H2 type	1.6e-05	26.8	5	304-326
862	mRNA_cap_enzyme	mRNA capping enzyme, catalytic domain	0.56	0.5	2	329-344
862	XPA_N	XPA protein N-terminal	0.78	5.1	4	329-341
862	zf-C2H2	Zinc finger, C2H2 type	5.4e-07	32.7	6	332-354
862	TFIIS	Transcription factor S-II (TFIIS)	0.29	6.5	5	360-370
862	zf-C2H2	Zinc finger, C2H2 type	1.1e-06	31.5	7	360-382
862	XPA_N	XPA protein N-terminal	0.13	7.8	6	385-397
862	TFIIS	Transcription factor S-II (TFIIS)	0.57	5.5	6	388-398
862	zf-C2H2	Zinc finger, C2H2 type	9.2e-07	31.8	8	388-410
862	XPA_N	XPA protein N-terminal	0.97	4.8	7	413-425
862	TFIIS	Transcription factor S-II (TFIIS)	0.14	7.6	7	416-426
862	zf-C2H2	Zinc finger, C2H2 type	4.4e-06	29.1	9	416-438
862	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.38	3.3	1	428-449
862	zf-C2H2	Zinc finger, C2H2 type	1.1e-06	31.5	10	444-466
862	TFIIS	Transcription factor S-II (TFIIS)	0.054	9.0	8	472-482
862	zf-C2H2	Zinc finger, C2H2 type	2.9e-07	33.8	11	472-494
862	zf-BED	BED zinc finger	0.64	4.8	3	477-495
862	DC1	1/2 472 487 .. 19 44	0.16	6.2	2	500-515
862	zf-C2H2	Zinc finger, C2H2 type	0.00082	19.9	12	500-523
863	Dor1	Dor1-like family	7e-203	684.1	1	197-553
863	bZIP	bZIP transcription factor	0.3	6.3	1	224-246
864	U1-C	U1 small nuclear ribonucleoprotein C	0.00024	16.9	1	2-51
864	zf-CCCH	Zinc finger C-x8-C-x5-C-x3-H type (an	2.2e-09	33.8	1	52-78
865	WD40	WD domain, G-beta repeat	4.2e-08	31.1	1	202-238
865	WD40	WD domain, G-beta repeat	0.54	6.3	2	282-307
866	Fels1	Fels-1 Propage Protein-like	0.61	5.8	1	361-376
867	aminotran_3	Aminotransferase class-III	1.5e-40	134.4	1	95-214
867	OATP_N	Organic Anion Transporter Polypeptide	0.81	4.0	1	240-258
867	aminotran_3	Aminotransferase class-III	8.9e-66	218.5	2	281-509
868	aminotran_3	Aminotransferase class-III	1.2e-09	31.3	1	52-111
868	OATP_N	Organic Anion Transporter Polypeptide	0.81	4.0	1	137-155
868	aminotran_3	Aminotransferase class-III	8.9e-66	218.5	2	178-406
869	trypsin	Trypsin	4.5e-71	220.5	1	63-289
870	Glycos_transf_1	Glycosyl transferases group 1	1.7e-17	64.4	1	144-239
872	MHYT	Bacterial signalling protein N termin	0.6	4.2	1	291-328
873	EGF	EGF-like domain	2.9e-07	28.9	1	7-43
873	laminin_EGF	Laminin EGF-like (Domains III and V)	1	4.3	1	21-43
873	EGF	EGF-like domain	9.2e-10	38.0	2	50-81
873	EGF	EGF-like domain	1.2e-07	30.3	3	88-119
873	EGF	EGF-like domain	2.7e-11	43.5	4	126-157
873	EGF	EGF-like domain	5e-11	42.5	5	168-199
873	DSL	Delta serrate ligand	0.32	5.2	3	190-199
873	EGF	EGF-like domain	0.0091	12.7	6	209-234

515

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
873	EGF	EGF-like domain	0.022	11.3	7	243-267
873	EGF	EGF-like domain	5e-09	35.3	8	280-311
873	EGF	EGF-like domain	1.3e-07	30.2	9	319-350
873	Cripto	Cripto growth factor	0.11	6.4	2	324-351
873	EGF	EGF-like domain	8.2e-11	41.8	10	358-389
873	Cripto	Cripto growth factor	0.00049	14.6	3	363-390
873	laminin_EGF	Laminin EGF-like (Domains III and V)	0.042	9.1	5	378-390
873	EGF	EGF-like domain	4.6e-08	31.8	11	396-427
873	laminin_EGF	Laminin EGF-like (Domains III and V)	0.25	6.4	6	416-427
873	sushi	Sushi domain (SCR repeat)	1.5e-06	28.7	1	433-486
873	EGF	EGF-like domain	8.7e-09	34.5	12	492-523
873	EGF	EGF-like domain	3.9e-09	35.7	13	530-561
873	EGF	EGF-like domain	1.2e-07	30.4	14	568-599
873	granulin	Granulin	1	3.6	2	596-608
873	EGF	EGF-like domain	2.9e-07	29.0	15	606-637
873	DSL	Delta serrate ligand	0.69	4.1	9	627-637
873	fn3	Fibronectin type III domain	1.3e-10	38.6	1	641-722
873	fn3	Fibronectin type III domain	8e-12	42.8	2	740-823
873	fn3	Fibronectin type III domain	1.2e-12	45.7	3	839-921
873	EGF	EGF-like domain	5.8e-10	38.7	16	1046-1077
873	Cripto	Cripto growth factor	0.047	7.7	5	1051-1078
875	AdoHcyase	S-adenosyl-L-homocysteine hydrolase	2.2e-68	222.4	1	81-217
875	AdoHcyase	S-adenosyl-L-homocysteine hydrolase	1.8e-55	180.1	2	218-507
875	AdoHcyase_NAD	S-adenosyl-L-homocysteine hydrolase,	2.2e-106	363.6	1	267-428
875	TrkA-N	TrkA-N domain	0.023	10.7	1	291-322
875	GlutR_NAD_bin d	Glutamyl-tRNA ^{Glu} reductase, NAD(P) bi	0.086	8.1	2	337-353
876	UQ_con	Ubiquitin-conjugating enzyme	0.0058	11.9	1	47-77
877	Prominin	Prominin	0	1616.6	1	18-823
877	SPDY	Domain of unknown function (DUF317)	0.15	6.5	1	80-93
877	DUF705	Protein of unknown function (DUF705)	0.98	1.9	1	555-565
878	fibrinogen_C	Fibrinogen beta and gamma chains, C-t	7.6e-56	190.6	1	146-382
879	fibrinogen_C	Fibrinogen beta and gamma chains, C-t	7.6e-56	190.6	1	146-382
880	fibrinogen_C	Fibrinogen beta and gamma chains, C-t	7.6e-56	190.6	1	146-382
881	DUF846	Eukaryotic protein of unknown functio	0.094	4.8	1	83-113
882	DUF381	Domain of unknown function (DUF381)	0.48	4.4	1	29-35
883	Trp_Tyr_perm	Tryptophan/tyrosine permease family	0.0026	10.3	1	42-63
883	aa_permeases	Amino acid permease	8.4e-32	115.8	1	48-371
883	Pox_I5	Poxvirus protein I5	0.24	6.0	1	162-179
883	serine_carbpept	Serine carboxypeptidase	0.41	2.3	1	378-398
884	pkinase	Protein kinase domain	6.3e-09	32.0	1	100-150
884	CtsR	Firmicute transcriptional repressor o	0.61	3.9	1	146-157
884	pkinase	Protein kinase domain	1.3e-07	27.2	2	151-181
884	Pox_ser-thr_kin	Poxvirus serine/threonine protein kin	0.31	3.8	1	165-176
884	Herpes_UL3	Herpesvirus UL3 protein	0.72	4.0	1	338-383
884	pkinase	Protein kinase domain	0.00084	13.7	3	444-495
884	pkinase	Protein kinase domain	2.1e-05	19.4	4	604-659
885	lectin_c	Lectin C-type domain	9.9e-10	40.5	1	47-107

516

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
886	spectrin	Spectrin repeat	0.4	5.5	1	1042-1095
887	spectrin	Spectrin repeat	0.4	5.5	1	1042-1095
888	Peptidase_M20	Peptidase family M20/M25/M40	5.2e-24	86.6	1	55-295
889	sugar_tr	Sugar (and other) transporter	0.11	5.5	1	47-103
889	Octopine_DH	NAD/NADP octopine/nopaline dehydrogen	0.26	4.6	1	153-169
889	sugar_tr	Sugar (and other) transporter	3.7e-08	28.6	2	201-335
890	T4_deiodinase	Iodothyronine deiodinase	0.37	4.0	1	168-179
891	ig	Immunoglobulin domain	8.5e-07	28.3	1	55-127
891	denso_VP4	Capsid protein VP4	0.38	2.7	1	57-69
892	bromodomain	Bromodomain	9.5e-45	158.8	1	63-152
892	bromodomain	Bromodomain	3e-40	143.5	2	356-445
892	Alpha_adaptin_C	Alpha adaptin AP2, C-terminal domain	0.48	2.6	1	395-407
892	Phage_X	Phage X family	0.97	3.7	1	438-469
892	eIF3c_N	Eukaryotic translation initiation fac	0.51	1.2	1	473-559
892	Vitellogenin_N	Lipoprotein amino terminal region	0.61	1.5	1	484-539
892	Herpes_U44	Herpes virus U44 protein	0.47	3.1	1	515-529
892	MAGP	Microfibril-associated glycoprotein (0.82	2.7	1	919-958
893	Pox_A_type_inc	Viral A-type inclusion protein repeat	0.23	7.6	1	197-216
893	OLF	Olfactomedin-like domain	4.6e-121	412.4	1	220-470
893	Phage_X	Phage X family	0.57	4.5	1	362-389
893	Peptidase_M10_N	Matrix metalloprotease, N-terminal do	0.86	2.1	1	373-383
893	FeThRed_B	Ferredoxin thioredoxin reductase cata	0.96	2.3	1	377-393
894	kazal	Kazal-type serine protease inhibitor	1.7e-10	44.0	1	88-132
894	efhand	EF hand	2.2e-05	23.3	1	178-206
894	ig	Immunoglobulin domain	6.4e-06	25.0	1	262-322
894	ig	Immunoglobulin domain	2e-09	38.2	2	354-414
894	SsgA	Streptomyces sporulation and cell div	0.35	5.9	1	541-549
895	aminotran_1_2	Aminotransferase class I and II	7.5e-20	71.8	1	81-257
895	DegT_DnrJ_EryC1	DegT/DnrJ/EryC1/StrS aminotransferase	1	2.4	1	158-178
895	TPP_enzymes_C	Thiamine pyrophosphate enzyme, C-term	0.35	3.3	1	258-279
896	LIM	LIM domain	9.9e-09	32.9	1	24-80
896	LIM	LIM domain	2e-13	49.7	2	83-134
896	LIM	LIM domain	5.3e-19	69.5	3	153-209
896	DUF866	Eukaryotic protein of unknown functio	0.035	7.5	1	178-199
896	LIM	LIM domain	7.5e-07	26.3	4	212-253
896	VHP	Villin headpiece domain	4.6e-25	77.5	1	538-573
897	LytTR	LytTr DNA-binding domain	0.051	9.5	1	14-49
897	COX4	Cytochrome c oxidase subunit IV	0.61	4.7	1	188-207
897	pkinase	Protein kinase domain	2.9e-102	349.9	1	356-613
897	TMP	TMP repeat	0.37	8.0	1	579-589
898	DCX	Doublecortin	1.4e-12	44.7	1	130-194
898	LytTR	LytTr DNA-binding domain	0.051	9.5	1	201-236
898	COX4	Cytochrome c oxidase subunit IV	0.61	4.7	1	375-394
898	pkinase	Protein kinase domain	2.9e-102	349.9	1	543-800
898	TMP	TMP repeat	0.37	8.0	1	766-776

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
899	glutaredoxin	Glutaredoxin	0.0052	12.1	1	101-154
899	GST_N	Glutathione S-transferase, N-terminal	0.053	9.4	1	102-152
899	ArsC	ArsC family	0.59	4.8	1	105-131
899	GST_C	Glutathione S-transferase, C-terminal	0.00013	17.6	1	278-370
899	UL21	Herpesvirus UL21	0.98	0.3	1	301-329
900	Collagen	Collagen triple helix repeat (20 copi	2.4e-05	22.3	1	27-60
900	Collagen	Collagen triple helix repeat (20 copi	1.5e-07	30.6	2	61-106
900	C1q	C1q domain	2.9e-72	250.2	1	116-241
900	TOBE	TOBE domain	0.5	6.3	1	207-226
901	Herpes_BMRF2	Herpesvirus BMRF2 protein	0.042	7.2	1	8-26
902	BRCT	BRCA1 C Terminus (BRCT) domain	5.9e-09	31.4	1	10-93
902	BRCT	BRCA1 C Terminus (BRCT) domain	1.4e-25	87.3	2	96-183
902	Sec6	Exocyst complex component Sec6	0.71	2.3	1	367-395
902	BRCT	BRCA1 C Terminus (BRCT) domain	7.8e-18	61.3	3	479-570
902	BRCT	BRCA1 C Terminus (BRCT) domain	5.7e-19	65.1	4	579-652
902	BRCT	BRCA1 C Terminus (BRCT) domain	2.3e-18	63.0	5	737-823
902	RinB	Transcriptional activator RinB	0.33	5.4	1	796-847
902	BRCT	BRCA1 C Terminus (BRCT) domain	0.028	9.0	6	846-881
902	Phage_Coat_A	Phage Coat Protein A	0.82	3.9	1	924-936
903	BRCT	BRCA1 C Terminus (BRCT) domain	5.9e-09	31.4	1	10-93
904	Phage_X	Phage X family	0.71	4.2	1	16-41
904	2OG-FeII_Oxy	2OG-Fe(II) oxygenase superfamily	0.27	6.0	1	195-273
905	LRR	Leucine Rich Repeat	0.0001	18.6	1	4-27
905	LRRCT	Leucine rich repeat C-terminal domain	4.3e-13	41.2	1	37-83
905	UPF0118	Domain of unknown function DUF20	1	2.9	1	211-234
906	ig	Immunoglobulin domain	7.9e-06	24.7	1	25-79
906	COX17	Cytochrome C oxidase copper chaperone	0.68	3.6	1	182-195
907	TB2_DP1_HVA22	TB2/DP1, HVA22 family	3.8e-34	123.6	1	3-96
907	ELM2	ELM2 domain	0.53	5.2	1	99-124
908	LRRNT	Leucine rich repeat N-terminal domain	0.00068	15.2	1	23-49
908	LRR	Leucine Rich Repeat	8.7e-05	18.9	1	51-74
908	Sal_vir_VRP3	Salmonella virulence-associated 28kDa	1	3.8	1	64-88
908	LRR	Leucine Rich Repeat	0.00012	18.4	2	75-98
908	LRR	Leucine Rich Repeat	0.0034	13.5	3	99-122
908	LRR	Leucine Rich Repeat	9.9e-06	22.1	4	123-146
908	LRRCT	Leucine rich repeat C-terminal domain	2.3e-15	48.2	1	156-208
908	ig	Immunoglobulin domain	1.3e-08	35.1	1	224-283
908	ig	Immunoglobulin domain	3.8e-09	37.1	2	320-376
908	ig	Immunoglobulin domain	0.00083	17.1	3	416-472
908	BON	Transport-associated domain	0.14	7.1	1	477-489
908	ig	Immunoglobulin domain	2.8e-08	33.9	4	533-590
908	pec_lyase_N	Pectate lyase, N terminus	0.19	3.9	1	670-676
908	An_peroxidase	Animal haem peroxidase	1.1e-193	653.6	1	770-1309
908	PAL	Phenylalanine and histidine ammonia-l	0.53	2.6	1	1037-1054
908	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	1101-1109
908	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	1194-1211
908	PetG	Cytochrome B6-F complex subunit 5	0.51	5.7	1	1245-1278

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
908	DUF978	Bacterial protein of unknown function	0.67	3.8	1	1257-1270
908	TILa	TILa domain	0.00018	16.9	1	1438-1477
908	PSP94	Beta-microseminoprotein (PSP-94)	0.11	8.0	1	1439-1470
908	vwc	von Willebrand factor type C domain	2e-10	38.0	1	1439-1494
909	LRRNT	Leucine rich repeat N-terminal domain	0.00068	15.2	1	54-80
909	LRR	Leucine Rich Repeat	8.7e-05	18.9	1	82-105
909	Sal_vir_VRP3	Salmonella virulence-associated 28kDa	1	3.8	1	95-119
909	LRR	Leucine Rich Repeat	0.00012	18.4	2	106-129
909	LRR	Leucine Rich Repeat	0.0034	13.5	3	130-153
909	LRR	Leucine Rich Repeat	9.9e-06	22.1	4	154-177
909	LRRCT	Leucine rich repeat C-terminal domain	2.3e-15	48.2	1	187-239
909	ig	Immunoglobulin domain	1.3e-08	35.1	1	255-314
909	ig	Immunoglobulin domain	3.8e-09	37.1	2	351-407
909	ig	Immunoglobulin domain	0.00083	17.1	3	447-503
909	BON	Transport-associated domain	0.14	7.1	1	508-520
909	ig	Immunoglobulin domain	2.8e-08	33.9	4	564-621
909	pec_lyase_N	Pectate lyase, N terminus	0.19	3.9	1	701-707
909	An_peroxidase	Animal haem peroxidase	1.1e-193	653.6	1	801-1340
909	PAL	Phenylalanine and histidine ammonia-l	0.53	2.6	1	1068-1085
909	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	1132-1140
909	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	1225-1242
909	PetG	Cytochrome B6-F complex subunit 5	0.51	5.7	1	1276-1309
909	DUF978	Bacterial protein of unknown function	0.67	3.8	1	1288-1301
909	TILa	TILa domain	0.00018	16.9	1	1469-1508
909	PSP94	Beta-microseminoprotein (PSP-94)	0.11	8.0	1	1470-1501
909	vwc	von Willebrand factor type C domain	2e-10	38.0	1	1470-1525
910	LRRNT	Leucine rich repeat N-terminal domain	0.00068	15.2	1	23-49
910	LRR	Leucine Rich Repeat	8.7e-05	18.9	1	51-74
910	LRR	Leucine Rich Repeat	0.00032	17.0	2	75-98
910	LRR	Leucine Rich Repeat	0.025	10.6	3	99-122
910	LRR	Leucine Rich Repeat	0.00069	15.8	4	123-146
910	ig	Immunoglobulin domain	1.3e-08	35.1	1	201-260
910	ig	Immunoglobulin domain	3.8e-09	37.1	2	297-353
910	ig	Immunoglobulin domain	0.00083	17.1	3	393-449
910	BON	Transport-associated domain	0.14	7.1	1	454-466
910	ig	Immunoglobulin domain	0.47	6.8	4	514-532
910	An_peroxidase	Animal haem peroxidase	1.1e-193	653.6	1	663-1202
910	PAL	Phenylalanine and histidine ammonia-l	0.53	2.6	1	930-947
910	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	994-1002

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
910	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	1087-1104
910	PetG	Cytochrome B6-F complex subunit 5	0.51	5.7	1	1138-1171
910	DUF978	Bacterial protein of unknown function	0.67	3.8	1	1150-1163
910	TILa	TILa domain	0.00018	16.9	1	1331-1370
910	PSP94	Beta-microseminoprotein (PSP-94)	0.11	8.0	1	1332-1363
910	vwc	von Willebrand factor type C domain	2e-10	38.0	1	1332-1387
911	EGF	EGF-like domain	0.059	9.8	2	47-59
911	EGF	EGF-like domain	0.0036	14.2	3	85-99
911	EGF	EGF-like domain	4.9e-08	31.7	4	106-134
911	EGF	EGF-like domain	4.2e-10	39.2	5	172-203
911	EGF	EGF-like domain	0.00083	16.5	6	210-245
911	laminin_EGF	Laminin EGF-like (Domains III and V)	0.014	10.8	3	216-247
911	laminin_G	Laminin G domain	0.0021	12.5	1	275-335
911	laminin_G	Laminin G domain	0.018	9.3	2	386-401
911	DUF604	Protein of unknown function, DUF604	0.84	2.9	1	390-412
911	laminin_G	Laminin G domain	0.22	5.5	3	483-541
911	EGF	EGF-like domain	9.9e-11	41.5	7	574-605
911	EGF	EGF-like domain	0.43	6.7	8	611-632
911	DUF1067	Protein of unknown function (DUF1067)	0.79	3.0	1	614-628
911	laminin_G	Laminin G domain	1.9e-05	19.6	4	663-728
911	Melibiose	Melibiose	0.9	2.3	1	740-755
911	laminin_G	Laminin G domain	0.075	7.2	5	773-788
911	EGF	EGF-like domain	2.2e-09	36.6	9	823-854
911	DSL	Delta serrate ligand	0.44	4.8	2	844-854
911	EGF	EGF-like domain	6.4e-06	24.1	10	861-892
911	EGF	EGF-like domain	0.71	5.9	11	901-933
911	DSL	Delta serrate ligand	0.67	4.2	4	923-933
911	EGF	EGF-like domain	3e-06	25.3	12	940-971
913	Omega-atracotox	Omega-atracotoxin	0.43	3.7	1	24-44
913	M	M protein repeat	0.28	8.8	1	146-166
913	UPF0137	Uncharacterised protein family (UPF01	0.04	7.4	1	322-347
914	RIIa	Regulatory subunit of type II PKA R-s	1e-14	54.8	1	25-62
914	SURF6	Surfeit locus protein 6	0.027	7.2	1	42-113
914	cNMP_binding	Cyclic nucleotide-binding domain	7.2e-31	112.5	1	152-240
914	RNA_pol_Rpb2_4	RNA polymerase Rpb2, domain 4	0.28	6.2	1	184-191
914	cNMP_binding	Cyclic nucleotide-binding domain	9.4e-32	115.7	2	270-364
914	Methyltransf_1	6-O-methylguanine DNA methyltransferase	0.64	4.3	1	325-337
915	DIL	DIL domain	1.8e-40	144.6	1	214-323
915	PDZ	PDZ domain (Also known as DHR or GLGF	1.7e-14	52.8	1	555-639
916	PLAT	PLAT/LH2 domain	9.8e-32	109.3	1	2-111
916	lipxygenase	Lipxygenase	3.9e-194	655.1	1	121-647
916	DUF181	Uncharacterized ACR, COG1944	0.81	2.4	1	247-258
916	PG_binding_1	Putative peptidoglycan binding domain	0.5	5.6	1	420-436

520

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
916	Dus	Dihydrouridine synthase (Dus)	0.18	4.6	1	604-647
917	PLAT	PLAT/LH2 domain	9.8e-32	109.3	1	2-111
917	lipoxygenase	Lipoxygenase	1.9e-48	164.1	1	112-293
917	DUF181	Uncharacterized ACR, COG1944	0.81	2.4	1	220-231
918	PLAT	PLAT/LH2 domain	9.8e-32	109.3	1	2-111
918	lipoxygenase	Lipoxygenase	2.7e-57	194.3	1	121-322
918	DUF181	Uncharacterized ACR, COG1944	0.81	2.4	1	249-260
920	TFIIS	Transcription factor S-II (TFIIS)	1	4.6	1	5-15
920	DUF536	Protein of unknown function, DUF536	0.19	7.9	1	214-251
920	FCH	Fes/CIP4 homology domain	0.5	5.6	1	259-278
926	DS	Deoxyhypusine synthase	0.53	2.5	1	21-36
926	SH3BP5	SH3 domain-binding protein 5 (SH3BP5)	0.097	6.5	1	82-102
926	Aa_trans	Transmembrane amino acid transporter	3.5e-139	472.5	1	114-517
926	Herpes_U47	Herpesvirus glycoprotein U47	0.69	1.1	1	141-158
926	Omega-atracotox	Omega-atracotoxin	0.35	4.0	1	168-184
926	DUF588	Domain of unknown function (DUF588)	0.58	5.1	1	425-444
926	GSPII_F	Bacterial type II secretion system pr	0.46	3.6	1	438-455
926	FtsX	Predicted permease	0.35	5.4	1	454-523
927	EGF	EGF-like domain	0.024	11.2	1	42-57
927	EGF	EGF-like domain	1.3e-06	26.6	2	60-88
927	EGF	EGF-like domain	1.2e-09	37.5	3	95-128
927	Cripto	Cripto growth factor	0.86	3.4	1	101-132
927	laminin_EGF	1/5 32 60 .. 2 43	0.025	9.9	2	106-130
927	EGF	EGF-like domain	5.5e-07	27.9	4	135-171
927	EGF	EGF-like domain	1e-10	41.4	5	178-209
927	EB	EB module	0.26	5.4	1	183-209
927	EGF	EGF-like domain	5e-08	31.7	6	216-247
927	DUF990	Protein of unknown function (DUF990)	0.23	5.3	1	302-336
927	MARVEL	Membrane-associating domain	0.15	5.8	1	305-333
927	PAP2	PAP2 superfamily	0.88	3.7	1	311-334
927	Colicin_V	Colicin V production protein	0.98	3.5	1	315-336
928	Ornatin	Ornatin	0.59	4.7	1	125-132
928	PP1_inhibitor	PKC-activated protein phosphatase-1 i	0.78	2.2	1	423-439
929	ank	Ankyrin repeat	0.011	12.7	2	142-167
930	LRRNT	Leucine rich repeat N-terminal domain	0.39	6.0	1	66-86
930	DUF6	Integral membrane protein DUF6	0.00023	18.9	1	86-129
930	DUF6	Integral membrane protein DUF6	7e-05	20.9	2	180-277
931	endotoxin	delta endotoxin	0.85	2.3	1	203-220
932	Lipoprotein_8	Hypothetical lipoprotein (MG045 famil	0.7	1.1	1	65-79
933	Peptidase_M24	metallopeptidase family M24	2.2e-70	244.0	1	88-326
933	DUF120	Domain of unknown function DUF120	0.089	7.1	1	169-180
934	Neurexophilin	Neurexophilin	2e-258	804.9	1	3-308
934	NnrS	NnrS protein	0.47	3.0	1	8-21
938	L27	L27 domain	7.3e-19	69.4	1	13-68
938	Not3	Not1 N-terminal domain, CCR4-Not comp	0.95	2.9	1	54-77
938	PDZ	PDZ domain (Also known as DHR or GLGF	8.1e-22	78.5	1	93-172
938	CDC50	LEM3 (ligand-effect modulator 3) fami	1	2.1	1	159-174
938	DUF100	Protein of unknown function DUF100	0.2	4.1	1	175-188
939	DIE2_ALG10	DIE2/ALG10 family	7.6e-72	248.9	1	28-146

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
939	DUF718	Protein of unknown function (DUF718)	0.64	4.4	1	36-43
939	Gemini_mov	Geminivirus putative movement protein	0.42	4.6	1	101-115
940	rrm	RNA recognition motif. (a.k.a. RRM, R	1.3e-09	36.2	1	61-128
940	RbsD_FucU	RbsD / FucU transport protein family	0.53	3.4	1	123-147
940	HemX	HemX	0.37	3.5	1	142-173
940	rrm	RNA recognition motif. (a.k.a. RRM, R	4.6e-13	48.6	2	186-253
940	rrm	RNA recognition motif. (a.k.a. RRM, R	4.3e-13	48.7	3	339-406
940	rrm	RNA recognition motif. (a.k.a. RRM, R	1.4e-06	25.5	4	456-524
941	C_tripleX	Cysteine rich repeat	2e-05	17.8	1	60-77
941	Bowman-Birk_leg	Bowman-Birk serine protease inhibitor	1	4.0	1	69-84
941	laminin_EGF	Laminin EGF-like (Domains III and V)	0.32	6.1	1	81-94
941	EGF	EGF-like domain	8.7e-06	23.6	2	99-127
941	TIL	Trypsin Inhibitor like cysteine rich	0.0035	11.0	1	118-139
941	EGF	EGF-like domain	7.5e-05	20.2	3	139-173
941	TIL	Trypsin Inhibitor like cysteine rich	0.26	5.1	2	152-179
941	toxin_5	Scorpion short toxin	0.34	4.4	1	154-159
941	EGF	EGF-like domain	4.4e-05	21.1	4	179-212
941	EGF	EGF-like domain	9.7e-09	34.3	5	224-259
941	MAM	MAM domain	3.5e-41	147.0	1	403-547
942	C_tripleX	Cysteine rich repeat	2e-05	17.8	1	65-82
942	Bowman-Birk_leg	Bowman-Birk serine protease inhibitor	1	4.0	1	74-89
942	laminin_EGF	Laminin EGF-like (Domains III and V)	0.32	6.1	1	86-99
942	EGF	EGF-like domain	8.7e-06	23.6	2	104-132
942	TIL	Trypsin Inhibitor like cysteine rich	0.0035	11.0	1	123-144
942	EGF	EGF-like domain	7.5e-05	20.2	3	144-178
942	TIL	Trypsin Inhibitor like cysteine rich	0.26	5.1	2	157-184
942	toxin_5	Scorpion short toxin	0.34	4.4	1	159-164
942	EGF	EGF-like domain	4.4e-05	21.1	4	184-217
942	EGF	EGF-like domain	9.7e-09	34.3	5	229-264
942	MAM	MAM domain	3.5e-41	147.0	1	408-552
943	PHD	PHD-finger	3.4e-14	45.7	1	85-128
943	bromodomain	Bromodomain	5.4e-12	44.0	1	149-235
943	PHD	PHD-finger	0.61	3.9	2	260-272
943	PWWP	PWWP domain	6.3e-10	36.2	1	269-312
943	GatB	PET112 family, C terminal region	0.64	5.1	1	288-303
943	TH1	TH1 protein	0.91	0.2	1	640-653
943	SP2	Structural protein 2	0.42	1.1	1	904-922
943	zf-B_box	B-box zinc finger	0.12	9.1	1	974-989
943	zf-MYND	MYND finger	5.3e-11	35.7	1	977-1011
944	PHD	PHD-finger	3.4e-14	45.7	1	85-128
944	bromodomain	Bromodomain	5.4e-12	44.0	1	149-235
944	PHD	PHD-finger	0.61	3.9	2	260-272
944	PWWP	PWWP domain	6.3e-10	36.2	1	269-312
944	GatB	PET112 family, C terminal region	0.64	5.1	1	288-303
944	TH1	TH1 protein	0.91	0.2	1	640-653
945	PHD	PHD-finger	3.4e-14	45.7	1	85-128
945	bromodomain	Bromodomain	5.4e-12	44.0	1	149-235
945	PHD	PHD-finger	0.61	3.9	2	260-272
945	PWWP	PWWP domain	6.3e-10	36.2	1	269-312
945	GatB	PET112 family, C terminal region	0.64	5.1	1	288-303

522

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
945	TH1	TH1 protein	0.91	0.2	1	640-653
945	SP2	Structural protein 2	0.42	1.1	1	950-968
945	zf-B_box	B-box zinc finger	0.12	9.1	1	1020-1035
945	zf-MYND	MYND finger	5.3e-11	35.7	1	1023-1057
946	PHD	PHD-finger	3.4e-14	45.7	1	90-133
946	bromodomain	Bromodomain	5.4e-12	44.0	1	154-240
946	PHD	PHD-finger	0.61	3.9	2	265-277
946	PWWP	PWWP domain	6.3e-10	36.2	1	274-317
946	GatB	PET112 family, C terminal region	0.64	5.1	1	293-308
946	TH1	TH1 protein	0.91	0.2	1	645-658
946	SP2	Structural protein 2	0.42	1.1	1	955-973
946	zf-B_box	B-box zinc finger	0.12	9.1	1	1025-1040
946	zf-MYND	MYND finger	5.3e-11	35.7	1	1028-1062
947	Urotensin_II	Urotensin II	0.36	5.4	1	362-372
947	fn2	Fibronectin type II domain	0.55	3.5	1	363-371
950	Terminase_5	Putative ATPase subunit of terminase	0.87	0.7	1	7-20
950	ion_trans	Ion transport protein	3.9e-08	29.8	1	345-518
950	SirB	Invasion gene expression up-regulator	0.2	6.0	1	350-366
950	Pept_C1-like	Peptidase C1-like family	0.88	1.2	1	549-569
950	BK_channel_a	Calcium-activated BK potassium channel	5.1e-07	22.5	1	598-702
950	zf-CHC2	CHC2 zinc finger	0.76	4.9	1	739-769
950	Alpha_adaptin_C	Alpha adaptin AP2, C-terminal domain	0.31	3.1	1	894-900
950	CPSase_L_D3	Carbamoyl-phosphate synthetase large	0.72	1.1	1	1086-1098
950	BK_channel_a	Calcium-activated BK potassium channel	0.029	5.8	2	1132-1171
951	Pep_M12B_prop ep	Reprolysin family propeptide	3.2e-37	116.5	1	80-198
951	Reprolysin	Reprolysin (M12B) family zinc metallo	1.1e-88	304.8	1	210-409
951	Fragilysin	Fragilysin metalloproteinase (M10C) en	0.28	3.8	1	342-355
951	Peptidase_M46	Pregnancy-associated plasma protein-A	0.056	5.5	1	345-355
951	disintegrin	Disintegrin	1.7e-39	134.2	1	426-501
951	EGF	EGF-like domain	0.95	5.4	1	631-654
953	ank	Ankyrin repeat	4.4e-06	24.9	1	151-179
953	ank	Ankyrin repeat	6.9e-09	35.0	2	183-215
953	ank	Ankyrin repeat	0.15	8.6	3	216-248
953	ank	Ankyrin repeat	9.7e-10	38.0	4	250-282
953	ank	Ankyrin repeat	0.00014	19.5	5	283-328
953	LolA	Outer membrane lipoprotein carrier pr	1	3.0	1	309-332
953	ank	Ankyrin repeat	3.8e-08	32.3	6	329-361
953	ank	Ankyrin repeat	0.49	6.8	7	362-394
954	interferon	Interferon alpha/beta domain	7.5e-42	144.5	1	16-105
955	ShTK	ShTK domain	0.46	4.9	1	67-74
955	NADHdh	NADH dehydrogenase	0.84	3.4	1	123-142
956	adh_short	short chain dehydrogenase	7.6e-27	92.5	1	31-137
956	sodcu	Copper/zinc superoxide dismutase (SOD)	0.059	5.9	1	70-87
956	Pex14_N	Peroxisomal membrane anchor protein (0.21	5.0	1	95-105

523

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
956	CitF	Citrate lyase, alpha subunit (CitF)	0.99	1.7	1	124-133
956	adh_short	short chain dehydrogenase	6.2e-11	37.7	2	138-188
957	thioered	Thioredoxin	1	4.4	1	69-96
957	Evr1_Alr	Erv1 / Alr family	5.1e-17	62.5	1	354-441
957	GAF	GAF domain	0.48	5.6	1	380-401
957	TFIIS	Transcription factor S-II (TFIIS)	0.14	7.6	1	394-406
958	acid_phosphat	Histidine acid phosphatase	6.4e-36	125.6	1	32-179
958	acid_phosphat	Histidine acid phosphatase	8.5e-24	83.0	2	205-381
958	NicO	High-affinity nickel-transport protei	0.99	2.9	1	398-416
959	serpin	Serpin (serine protease inhibitor)	8.5e-197	663.9	1	1-329
960	serpin	Serpin (serine protease inhibitor)	6.8e-87	295.8	1	45-191
960	serpin	Serpin (serine protease inhibitor)	1.7e-116	396.7	2	192-397
961	serpin	Serpin (serine protease inhibitor)	1.6e-63	216.1	1	45-158
961	serpin	Serpin (serine protease inhibitor)	5e-139	472.0	2	159-397
962	serpin	Serpin (serine protease inhibitor)	4.5e-151	512.0	1	45-300
962	Molydop_bindin g	Molydopterin dinucleotide binding dom	0.89	4.1	1	289-309
962	serpin	Serpin (serine protease inhibitor)	4.9e-56	190.5	2	301-397
963	OprB	Carbohydrate-selective porin, OprB fa	0.047	6.5	1	16-33
963	Alliinase_C	Allinase, C-terminal domain	0.63	4.1	1	45-58
963	Adeno_E1A	Early E1A protein	0.33	2.4	1	237-251
964	Pep_M12B_prop ep	Reprolysin family propeptide	2.4e-47	148.0	1	112-220
964	Reprolysin	Reprolysin (M12B) family zinc metallo	1.9e-96	330.6	1	232-426
964	Astacin	Astacin (Peptidase family M12A)	0.21	5.0	1	366-380
964	Phi_1	Phosphate-induced protein 1 conserved	0.51	3.3	1	414-426
964	disintegrin	Disintegrin	5.8e-23	78.5	1	444-517
964	CBM_10	Cellulose or protein binding domain	0.47	6.8	1	481-499
964	EGF	EGF-like domain	0.21	7.8	2	664-693
965	Uteroglobin	Uteroglobin family	6.6e-09	29.8	1	1-88
966	7tm_2	7 transmembrane receptor (Secretin fa	0.96	2.6	1	19-38
966	GDA1_CD39	GDA1/CD39 (nucleoside phosphatase) fa	2.2e-93	315.4	1	48-483
966	E1	Papillomavirus helicase	0.36	4.3	1	76-92
966	PLRV_ORF5	Potato leaf roll virus readthrough pr	0.72	1.6	1	143-161
966	Nicastrin	Nicastrin	0.65	1.6	1	146-171
966	DUF462	Protein of unknown function, DUF462	0.55	4.7	1	371-390
966	Adeno_E3B	Adenovirus E3B protein	0.7	3.6	1	495-502
967	C1q	C1q domain	6.1e-44	156.1	1	73-202
968	Ornatin	Ornatin	0.55	4.8	1	99-106
969	Ornatin	Ornatin	0.55	4.8	1	134-141
969	Spo7	Spo7-like protein	1	1.5	1	405-417
969	MARVEL	Membrane-associating domain	0.37	4.5	1	487-526
969	DUF202	Domain of unknown function DUF	0.23	5.7	1	493-518
970	ig	Immunoglobulin domain	0.0038	14.6	1	41-124
970	ig	Immunoglobulin domain	0.00023	19.2	2	163-230
970	Gag_p30	Gag P30 core shell protein	3.6e-08	28.0	1	452-491
970	zf-CCHC	Zinc knuckle	8.8e-07	27.8	1	523-540
971	Prefoldin	Prefoldin subunit	0.66	5.0	1	179-206
971	Seryl_tRNA_N	Seryl-tRNA synthetase N-terminal doma	0.92	5.7	1	179-196

524

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
971	Adeno_PIX	Adenovirus hexon-associated protein (0.12	6.6	1	181-203
971	pentaxin	Pentaxin family	2.3e-26	91.1	1	302-464
971	Avirulence	Xanthomonas avirulence protein, Avr/P	0.07	3.6	1	439-453
972	ArsA_ATPase	Anion-transporting ATPase	0.87	2.4	1	59-69
972	TSPN	Thrombospondin N-terminal -like domai	0.88	2.7	1	223-255
972	RHS_repeat	RHS Repeat	0.00085	15.6	2	239-266
972	RHS_repeat	RHS Repeat	6.6e-05	19.5	4	314-367
973	bZIP	bZIP transcription factor	0.00024	17.2	1	623-686
973	integrase_DNA	DNA binding domain of tn916 integrase	0.38	6.3	1	657-693
973	CarD_TRCF	CarD-like/TRCF domain	0.54	4.5	1	708-728
974	WD40	WD domain, G-beta repeat	0.05	9.9	1	2-27
974	DUF596	Protein of unknown function, DUF596	0.84	3.7	1	63-76
974	WD40	WD domain, G-beta repeat	0.29	7.2	3	76-109
974	denso_VP4	Capsid protein VP4	0.81	1.5	1	355-364
974	TPR	TPR Domain	0.1	9.1	1	742-767
974	Paramyxo_C	Paramyxovirus non-structural protein	0.74	2.8	1	784-800
974	Xylose_isom	Xylose isomerase	0.4	3.2	1	796-811
974	TPR	TPR Domain	0.083	9.4	2	962-990
974	U-box	U-box domain	0.036	6.5	1	1294-1308
975	cofilin_ADF	Cofilin/tropomyosin-type actin-bindin	0.97	4.0	1	6-18
975	Phage_CII	Bacteriophage CII protein	1	3.9	1	229-243
975	ion_trans	Ion transport protein	0.0048	11.5	1	247-408
975	Sarcolipin	Sarcolipin	0.56	5.3	1	362-390
976	cofilin_ADF	Cofilin/tropomyosin-type actin-bindin	0.97	4.0	1	6-18
976	Phage_CII	Bacteriophage CII protein	1	3.9	1	303-317
976	ion_trans	Ion transport protein	0.0048	11.5	1	321-482
976	Sarcolipin	Sarcolipin	0.56	5.3	1	436-464
977	zf-C2H2	Zinc finger, C2H2 type	0.083	11.8	1	4-27
977	zf-C2H2	Zinc finger, C2H2 type	0.00081	19.9	2	108-131
977	zf-C2H2	Zinc finger, C2H2 type	0.07	12.1	3	162-185
977	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.45	3.1	1	238-248
977	zf-C2H2	Zinc finger, C2H2 type	0.28	9.7	5	439-462
977	zf-C2H2	Zinc finger, C2H2 type	0.0026	17.9	7	600-623
977	zf-C2H2	Zinc finger, C2H2 type	0.047	12.8	9	886-908
977	zf-C2H2	Zinc finger, C2H2 type	0.66	8.2	11	1030-1053
977	zf-C2H2	Zinc finger, C2H2 type	0.025	13.9	14	1265-1288
977	adeno_fiber	Adenoviral fibre protein (knob domain	0.076	3.5	1	1349-1357
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.023	14.1	17	1470-1493
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.031	13.5	19	1577-1600
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.022	14.1	20	1660-1683
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.0044	16.9	23	1892-1914
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.41	9.0	24	1968-1990
977	DC1	DC1 domain	0.68	4.3	2	2049-

525

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						2064
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.0039	17.2	25	2051-2073
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.0014	18.9	26	2085-2107
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.0094	15.6	27	2114-2137
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.041	13.0	28	2143-2166
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.033	13.4	30	2280-2303
977	TFIID-31	Transcription initiation factor IID,	0.28	5.7	1	2300-2315
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.14	10.9	31	2314-2336
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.0018	18.6	32	2360-2382
977	zf-C2H2	16/34 1369 1392 .. 1 24	0.016	14.7	33	2388-2411
977	Histone_HNS	H-NS histone family	0.85	4.7	1	2423-2434
977	zf-C2H2	16/34 1369 1392 .. 1 24	3.6e-05	25.4	34	2474-2496
977	PdxA	Pyridoxal phosphate biosynthetic prot	0.41	4.2	1	2540-2561
980	IGFBP	Insulin-like growth factor binding pr	0.017	10.0	1	24-56
980	kazal	Kazal-type serine protease inhibitor	9.3e-07	29.4	1	71-117
980	trypsin	Trypsin	4.2e-24	74.5	1	167-326
980	LSM	LSM domain	0.27	7.4	1	186-209
980	DUF771	Domain of unknown function (DUF771)	0.21	5.2	1	307-322
980	PDZ	PDZ domain (Also known as DHR or GLGF	7.1e-14	50.6	1	332-427
981	asp	Eukaryotic aspartyl protease	6.6e-35	123.8	1	19-112
981	trans_reg_C	Transcriptional regulatory protein, C	0.019	11.1	1	27-55
981	asp	Eukaryotic aspartyl protease	1.8e-23	83.1	2	165-239
981	asp	Eukaryotic aspartyl protease	0.0003	14.7	3	240-268
981	asp	Eukaryotic aspartyl protease	1.7e-48	171.3	4	295-421
984	Zn_carbOpept	Zinc carboxypeptidase	1.2e-76	259.4	1	48-249
984	APC_basic	APC basic domain	0.53	2.7	1	279-292
985	Zn_carbOpept	Zinc carboxypeptidase	1.2e-76	259.4	1	48-249
985	APC_basic	APC basic domain	0.53	2.7	1	279-292
986	NifU_N	NifU-like N terminal domain	1.7e-80	277.6	1	34-160
987	SNF7	SNF7	6.6e-65	225.8	1	108-277
987	Glyco_tran_28_C	Glycosyltransferase family 28 C-termi	0.71	3.8	1	171-201
988	Rz1	Lipoprotein Rz1 precursor	0.92	4.2	1	1-35
988	UPAR_LY6	u-PAR/Ly-6 domain	6.4e-06	29.8	1	28-110
990	zf-C2H2	Zinc finger, C2H2 type	0.00035	21.4	1	53-78
990	zf-C2H2	Zinc finger, C2H2 type	0.012	15.2	2	87-114
990	zf-C2H2	Zinc finger, C2H2 type	0.0039	17.1	3	120-144
991	pkinase	Protein kinase domain	3.2e-90	309.9	1	20-312
991	Glyco_hydro_15	Glycosyl hydrolases family 15	0.18	4.4	1	472-522
992	Prefoldin	Prefoldin subunit	0.12	7.6	1	5-44
992	spectrin	Spectrin repeat	0.00067	15.0	1	59-121

526
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
992	spectrin	Spectrin repeat	5.4e-06	22.2	2	124-226
992	DUF16	Protein of unknown function DUF16	0.67	3.9	1	202-250
992	spectrin	Spectrin repeat	5.2e-07	25.7	3	229-340
992	GSPII_E_N	GSPII_E N-terminal domain	0.07	7.7	1	265-290
992	spectrin	Spectrin repeat	2.8e-05	19.8	4	343-449
992	TelA	Toxic anion resistance protein (TelA)	0.75	3.2	1	405-437
992	spectrin	Spectrin repeat	2e-06	23.7	5	452-538
992	spectrin	Spectrin repeat	3.1e-13	47.2	6	781-888
992	DCP2	Dcp2, box A domain	0.57	4.2	1	823-837
992	MutS_II	MutS domain II	0.91	3.5	1	840-869
992	SAA_proteins	Serum amyloid A protein	0.07	6.0	1	866-883
993	LysE	LysE type translocator	0.02	8.8	1	127-147
994	Collagen	Collagen triple helix repeat (20 copi	9.2e-07	27.7	1	76-118
994	C1q	C1q domain	8e-32	115.9	1	160-284
995	Allantoicase	Allantoicase repeat	2.1e-75	257.1	1	1-136
995	Allantoicase	Allantoicase repeat	6.6e-58	197.5	2	159-319
996	DNA_ligase_A_C	ATP dependent DNA ligase C terminal r	0.67	5.4	1	11-34
996	ig	Immunoglobulin domain	0.00019	19.5	1	37-151
996	ig	Immunoglobulin domain	0.15	8.7	2	182-243
996	ig	Immunoglobulin domain	0.0031	15.0	3	275-335
996	SK_channel	Calcium-activated SK potassium channe	0.035	7.1	1	363-383
997	PH	PH domain	2.4e-24	81.6	1	23-133
997	HS2ST	Heparan sulfate 2-O-sulfotransferase	0.27	4.4	1	140-162
997	LMP	LMP repeated region	0.0012	14.2	1	160-181
997	DUF603	Protein of unknown function, DUF603	0.04	6.4	1	173-187
997	Pox_A_type_inc	Viral A-type inclusion protein repeat	0.32	7.2	1	173-187
997	IQ	IQ calmodulin-binding motif	5e-05	20.1	1	206-226
997	RhoGEF	RhoGEF domain	1.2e-69	236.9	1	247-428
997	DUF674	Protein of unknown function (DUF674)	0.82	1.4	1	275-285
997	Stig1	Stigma-specific protein, Stig1	0.6	2.3	1	376-421
997	PH	PH domain	2.3e-13	45.3	2	460-588
997	RasGEFN	Guanine nucleotide exchange factor fo	1.1e-19	71.3	1	633-688
997	RasGEF	RasGEF domain	7.2e-89	305.4	1	999-1184
997	Adeno_terminal	Adenoviral DNA terminal protein	1	1.7	1	1175-1207
998	DUF630	Protein of unknown function (DUF630)	0.7	4.3	1	692-705
998	FGF	Fibroblast growth factor	0.37	4.4	1	728-743
998	tRNA-synt_2	tRNA synthetases class II (D, K and N	0.74	3.5	1	754-766
998	Omega-atracotox	Omega-atracotoxin	0.15	5.1	1	859-866
999	K_tetra	K+ channel tetramerisation domain	2e-34	121.3	1	26-114
999	BTB	BTB/POZ domain	0.0015	14.2	1	74-125
1000	PXA	PXA domain	0.01	10.2	1	84-104
1000	Vps52	Vps52 / Sac2 family	0	1099.3	1	94-601
1000	trp_syntA	Tryptophan synthase alpha chain	0.78	3.1	1	173-210
1000	DUF965	Bacterial protein of unknown function	0.33	4.5	1	285-298
1000	Vps53_N	Vps53-like, N-terminal	0.93	2.7	1	565-585
1001	PHD	PHD-finger	3.8e-06	20.3	1	1-24
1001	rubredoxin	Rubredoxin	0.55	5.9	1	14-28
1001	Orbi_NS3	Orbivirus NS3	0.83	2.8	1	435-458
1001	NosL	NosL	0.29	4.9	1	1297-

527

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						1321
1001	NAC	NAC domain	0.76	5.5	1	1343-1365
1001	DUF240	MG032/MG096/MG288 family 2	0.17	6.7	1	1369-1384
1002	RecR	RecR protein	0.97	6.3	1	104-118
1002	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	2e-09	26.7	1	108-147
1002	DC1	DC1 domain	0.045	7.9	1	184-213
1002	PHD	PHD-finger	6.5e-21	66.9	2	185-233
1002	zf-MYND	MYND finger	0.7	4.3	1	186-204
1002	rubredoxin	Rubredoxin	0.55	5.9	1	223-237
1002	Orbi_NS3	Orbivirus NS3	0.83	2.8	1	644-667
1002	NosL	NosL	0.29	4.9	1	1506-1530
1002	NAC	NAC domain	0.76	5.5	1	1552-1574
1002	DUF240	MG032/MG096/MG288 family 2	0.17	6.7	1	1578-1593
1003	Patched	Patched family	0.069	4.7	1	405-442
1003	ISAV_HA	Infectious salmon anaemia virus haema	0.23	3.2	1	716-738
1003	WD40	WD domain, G-beta repeat	0.00019	18.3	1	767-802
1003	WD40	WD domain, G-beta repeat	0.71	5.9	2	958-992
1003	WD40	WD domain, G-beta repeat	4.2e-05	20.6	3	1069-1104
1003	WD40	WD domain, G-beta repeat	4.1e-09	34.6	4	1109-1145
1003	WD40	WD domain, G-beta repeat	0.0012	15.6	5	1150-1185
1004	ZZ	Zinc finger, ZZ type	1e-12	48.2	1	3-48
1004	SoxD	Sarcosine oxidase, delta subunit fami	0.97	4.2	1	77-84
1004	zf-C2H2	Zinc finger, C2H2 type	0.00067	20.3	1	78-101
1004	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.3	3.6	1	93-113
1004	SPDY	Domain of unknown function (DUF317)	0.6	4.4	1	117-131
1004	Di19	Drought induced 19 protein (Di19)	0.00056	13.0	1	312-328
1006	C2	C2 domain	7e-08	28.1	1	189-259
1006	HHE	Domain of Unknown function	0.13	7.5	1	216-235
1006	C2	C2 domain	1.3e-18	64.3	2	304-394
1007	RmuC	RmuC family	0.79	3.1	1	4-34
1007	IBN_NT	Importin-beta N-terminal domain	2.1e-27	99.5	1	22-101
1007	Peripla_BP_like	Periplasmic binding proteins and suga	0.21	4.7	1	130-161
1008	Las1	Las1-like	1.6e-94	320.7	1	38-186
1008	MuDR	MuDR family transposase	0.17	5.5	1	214-246
1008	BAR	BAR domain	0.21	5.2	1	330-346
1008	Adeno_E1B_19K	Adenovirus E1B 19K protein / small t-	0.43	4.6	1	517-541
1008	META	Domain of unknown function (306)	0.7	5.7	1	615-648
1009	PH	PH domain	4e-18	61.1	1	136-227
1009	HrpF	HrpF protein	0.64	4.5	1	248-257
1009	ArfGap	Putative GTPase activating protein fo	4.8e-38	133.0	1	250-373
1009	ank	Ankyrin repeat	3.2e-05	21.8	1	411-446
1009	ank	Ankyrin repeat	0.00019	19.0	2	447-479
1009	DMRL_synthase	6,7-dimethyl-8-ribityllumazine syntha	0.35	5.0	1	479-494
1009	hormone	Somatotropin hormone family	0.5	1.6	1	545-561
1009	tubulin-binding	Tau and MAP protein, tubulin-binding	0.11	8.0	1	828-844

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1009	SH3	SH3 domain	5.8e-12	46.6	1	881-938
1010	Bromo_CP	Bromovirus coat protein	0.16	5.5	1	1-12
1011	ig	Immunoglobulin domain	0.004	14.6	1	27-45
1011	ig	Immunoglobulin domain	3.9e-05	22.1	2	80-148
1011	ig	Immunoglobulin domain	3.7e-10	40.9	3	183-242
1011	ig	Immunoglobulin domain	0.0018	15.9	4	281-342
1011	ig	Immunoglobulin domain	3.7e-08	33.4	5	379-440
1011	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	396-452
1011	OapA	Opacity-associated protein A	0.44	2.4	1	405-427
1011	ig	Immunoglobulin domain	0.0012	16.6	6	474-535
1011	ig	Immunoglobulin domain	7.7e-07	28.5	7	570-634
1013	denso_VP4	Capsid protein VP4	0.23	3.4	1	166-185
1015	efhand	EF hand	2.8e-08	33.9	1	29-57
1015	COX17	Cytochrome C oxidase copper chaperone	0.42	4.2	1	54-61
1015	efhand	EF hand	0.0033	15.3	2	65-93
1015	efhand	EF hand	8.5e-05	21.1	3	102-130
1015	PCRF	PCRF domain	0.43	6.1	1	129-145
1015	DUF21	Domain of unknown function DUF21	0.18	6.4	1	134-158
1015	efhand	EF hand	5e-09	36.7	4	138-166
1016	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	3.9e-74	256.4	1	2-279
1016	Flavodoxin_2	Flavodoxin-like fold	0.66	3.3	1	373-388
1016	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-05	19.1	2	403-444
1017	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1e-39	140.9	1	119-253
1017	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	6.8e-52	182.6	2	431-611
1017	Flavodoxin_2	Flavodoxin-like fold	0.66	3.3	1	705-720
1017	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-05	19.1	3	735-776
1018	7tm_1	7 transmembrane receptor (rhodopsin f	1.1e-88	264.6	1	87-350
1018	DUF395	YeeE/YedE family (DUF395)	0.94	4.7	1	188-205
1019	LRRNT	Leucine rich repeat N-terminal domain	0.12	7.7	1	42-56
1019	LRR	Leucine Rich Repeat	0.12	8.2	1	82-105
1019	LRR	Leucine Rich Repeat	0.0019	14.3	3	133-157
1019	LRR	Leucine Rich Repeat	0.013	11.6	4	158-181
1019	LRR	Leucine Rich Repeat	0.00023	17.5	5	182-205
1019	LRR	Leucine Rich Repeat	0.31	6.9	6	206-226
1019	LRR	Leucine Rich Repeat	0.22	7.4	8	251-272
1019	LRR	9/18 273 283 .. 1 11	0.00057	16.1	10	329-352
1019	LRR	9/18 273 283 .. 1 11	0.004	13.3	12	377-402
1019	LRR	9/18 273 283 .. 1 11	0.0013	14.9	13	403-426
1019	LRR	9/18 273 283 .. 1 11	0.27	7.1	14	427-439
1019	LRR	9/18 273 283 .. 1 11	0.16	7.9	15	463-484
1019	LRR	9/18 273 283 .. 1 11	0.8	5.5	16	486-510
1019	LRR	9/18 273 283 .. 1 11	0.035	10.1	17	537-558
1019	TIMELESS	Timeless protein	0.45	3.0	1	553-568
1019	LRR	9/18 273 283 .. 1 11	0.084	8.8	18	559-582
1020	AMP-binding	AMP-binding enzyme	4.5e-49	173.2	1	1-177
1020	RNA_pol_Rpc4	RNA polymerase III RPC4	0.62	4.2	1	189-199
1020	Phage_30_8	Phage GP30.8 protein	0.92	2.6	1	233-253

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1021	SKIP_SNW	SKIP/SNW domain	0.3	4.7	1	92-113
1021	cNMP_binding	Cyclic nucleotide-binding domain	0.55	5.2	1	102-132
1021	cytochrome_c	Cytochrome c	0.92	3.7	1	313-329
1021	cNMP_binding	Cyclic nucleotide-binding domain	1.5e-15	57.4	2	345-435
1021	RasGEFN	Guanine nucleotide exchange factor fo	0.00023	17.5	1	460-504
1021	Pseu_avirulence	Avirulence protein	0.91	1.9	1	491-504
1021	PDZ	PDZ domain (Also known as DHR or GLGF	5.2e-19	68.7	1	580-661
1021	RA	Ras association (RalGDS/AF-6) domain	2.6e-08	32.5	1	806-885
1021	RasGEF	RasGEF domain	2.7e-48	170.6	1	907-1092
1022	SKIP_SNW	SKIP/SNW domain	0.3	4.7	1	42-63
1022	cNMP_binding	Cyclic nucleotide-binding domain	0.55	5.2	1	52-82
1022	cytochrome_c	Cytochrome c	0.92	3.7	1	263-279
1022	cNMP_binding	Cyclic nucleotide-binding domain	1.5e-15	57.4	2	295-385
1022	RasGEFN	Guanine nucleotide exchange factor fo	0.00023	17.5	1	410-454
1022	Pseu_avirulence	Avirulence protein	0.91	1.9	1	441-454
1022	PDZ	PDZ domain (Also known as DHR or GLGF	5.2e-19	68.7	1	530-611
1022	RA	Ras association (RalGDS/AF-6) domain	2.6e-08	32.5	1	756-835
1022	RasGEF	RasGEF domain	2.7e-48	170.6	1	857-1042
1026	Ricin_B_lectin	QXW lectin repeat	0.14	8.3	1	134-161
1026	MCR_beta_N	Methyl-coenzyme M reductase beta subu	0.98	2.1	1	152-160
1026	Ricin_B_lectin	QXW lectin repeat	4.5e-07	28.1	2	196-225
1026	Ricin_B_lectin	QXW lectin repeat	0.0012	15.8	3	226-265
1027	SCF	Stem cell factor	2.9e-155	512.2	1	1-214
1027	FH2	Formin Homology 2 Domain	0.027	8.8	1	145-162
1027	Herpes_UL7	Herpesvirus UL7 like	0.072	7.6	1	176-215
1028	cadherin	Cadherin domain	3.4e-12	44.2	1	50-131
1028	cadherin	Cadherin domain	1.7e-22	80.1	2	155-250
1028	cadherin	Cadherin domain	6e-20	71.3	3	264-342
1028	cadherin	Cadherin domain	5.9e-21	74.8	4	379-452
1028	cadherin	Cadherin domain	0.0035	12.8	5	521-567
1029	cadherin	Cadherin domain	3.4e-12	44.2	1	50-131
1029	cadherin	Cadherin domain	1.7e-22	80.1	2	155-250
1029	cadherin	Cadherin domain	6e-20	71.3	3	264-342
1029	cadherin	Cadherin domain	1.8e-22	80.0	4	379-470
1029	cadherin	Cadherin domain	0.0035	12.8	5	483-529
1030	Troponin	Troponin	0.87	3.1	1	21-117
1030	Mycoplasma_MAA2	Mycoplasma arthritidis MAA2 repeat	0.65	3.7	1	518-527
1030	PH	PH domain	6.5e-14	47.1	1	522-624
1030	DUF1041	Domain of Unknown Function (DUF1041)	3.4e-79	273.2	1	738-950
1030	Allene_ox_cyc	Allene oxide cyclase	0.7	2.8	1	817-852
1031	Renal_dipeptase	Renal dipeptidase	1.9e-108	370.4	1	74-354
1031	Amidase_3	N-acetylmuramoyl-L-alanine amidase	0.76	3.8	1	222-234
1032	Trp_Tyr_perm	Tryptophan/tyrosine permease family	0.0026	10.3	1	42-63

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1032	aa_permeases	Amino acid permease	8.4e-32	115.8	1	48-371
1032	Pox_I5	Poxvirus protein I5	0.24	6.0	1	162-179
1032	serine_carbpept	Serine carboxypeptidase	0.41	2.3	1	378-398
1033	THF_DHG_CYH	Tetrahydrofolate dehydrogenase/cycloh	0.027	6.2	1	89-108
1033	THF_DHG_CYH	Tetrahydrofolate dehydrogenase/cycloh	6.1e-13	37.3	2	119-180
1033	THF_DHG_CYH C	Tetrahydrofolate dehydrogenase/cycloh	6.5e-07	25.7	1	182-229
1033	FTHFS	Formate--tetrahydrofolate ligase	0	1365. 1	1	360-979
1034	acid_phosphat	Histidine acid phosphatase	0.038	6.9	1	378-394
1034	FMN_red	NADPH-dependent FMN reductase	0.94	3.3	1	425-446
1034	acid_phosphat	Histidine acid phosphatase	0.02	7.9	2	512-581
1034	Ribosomal_L6	Ribosomal protein L6	0.21	7.2	1	760-800
1035	PDZ	PDZ domain (Also known as DHR or GLGF)	3.2e-14	51.8	1	47-111
1035	DUF62	Protein of unknown function DUF62	1	2.7	1	71-91
1035	AraC_binding	AraC-like ligand binding domain	0.99	3.9	1	139-198
1035	Armadillo_seg	Armadillo/beta-catenin-like repeat	0.97	5.6	1	170-187
1035	HCV_NS4a	Hepatitis C virus non-structural prot	0.057	8.8	1	319-348
1035	RasGAP	GTPase-activator protein for Ras-like	0.37	3.6	1	764-783
1035	RhoGEF	RhoGEF domain	1.7e-28	97.2	1	778-962
1035	SH2	SH2 domain	0.98	3.2	1	819-829
1035	PH	PH domain	4.2e-05	18.0	1	1006- 1119
1035	SelP_N	Selenoprotein P, N terminal region	0.25	3.7	1	1112- 1138
1037	PH	PH domain	6.7e-14	47.1	1	17-124
1037	efhand	EF hand	9.2e-05	21.0	1	138-166
1037	efhand	EF hand	0.0023	15.8	2	174-202
1037	PI-PLC-X	Phosphatidylinositol-specific phospho	5.9e-17	60.5	1	291-326
1038	DUF765	Circovirus protein of unknown functio	0.85	3.7	1	274-302
1039	ABG_transport	AbgT putative transporter family	0.81	1.2	1	13-26
1039	7tm_1	7 transmembrane receptor (rhodopsin f	7.4e-29	85.1	1	40-289
1039	HECT	HECT-domain (ubiquitin-transferase)	0.15	5.5	1	273-290
1040	TSPN	Thrombospondin N-terminal -like domai	1.4e-41	136.6	1	1-101
1040	TIL	Trypsin Inhibitor like cysteine rich	0.66	3.9	1	195-239
1040	EGF	EGF-like domain	0.0046	13.8	1	199-233
1040	Baculo_LEF-3	Nucleopolyhedrovirus late expression	0.0024	10.4	1	230-244
1040	EGF	EGF-like domain	0.51	6.4	2	239-269
1040	Mu-conotoxin	Mu-Conotoxin	0.63	5.2	1	283-304
1040	dickkopf_N	Dickkopf N-terminal cysteine-rich reg	0.94	3.5	1	292-299
1040	laminin_EGF	Laminin EGF-like (Domains III and V)	0.45	5.6	1	311-327
1040	EGF	EGF-like domain	4.3e-05	21.1	4	333-366
1040	tsp_3	Thrombospondin type 3 repeat	0.00027	16.9	1	405-417
1040	tsp_3	Thrombospondin type 3 repeat	0.032	10.2	2	418-433
1040	tsp_3	Thrombospondin type 3 repeat	0.0046	13.0	3	441-453
1040	tsp_3	Thrombospondin type 3 repeat	0.00087	15.3	4	464-476
1040	tsp_3	Thrombospondin type 3 repeat	0.023	10.7	5	477-492
1040	tsp_3	Thrombospondin type 3 repeat	0.00058	15.9	6	500-512
1040	tsp_3	Thrombospondin type 3 repeat	0.0033	13.4	7	523-535
1040	tsp_3	Thrombospondin type 3 repeat	0.0011	15.0	8	538-553
1040	tsp_3	Thrombospondin type 3 repeat	0.00057	15.9	9	561-573
1040	tsp_3	Thrombospondin type 3 repeat	0.0015	14.6	11	601-613

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1040	tsp_3	Thrombospondin type 3 repeat	0.03	10.3	12	614-629
1040	TSPC	Thrombospondin C-terminal region	7.1e-176	594.4	1	654-854
1040	Mnd1	Mnd1 family	0.68	3.4	1	853-861
1042	sodcu	Copper/zinc superoxide dismutase (SOD)	1	2.0	1	31-44
1042	DapB_C	Dihydrodipicolinate reductase, C-term	0.84	4.5	1	38-52
1042	PTR2	POT family	3e-55	193.1	1	103-335
1042	PTR2	POT family	1.5e-36	127.7	2	336-471
1042	Adeno_PIX	Adenovirus hexon-associated protein (0.76	3.8	1	493-508
1043	Drf_GBD	Diaphanous GTPase-binding Domain	1.7e-60	211.1	1	40-229
1043	DUF1000	Domain of Unknown Function (DUF1000)	0.79	3.3	1	141-157
1043	Drf_FH3	Diaphanous FH3 Domain	7.2e-71	244.6	1	231-437
1043	S-antigen	S-antigen protein	0.24	3.2	1	429-436
1043	bZIP	bZIP transcription factor	0.88	4.7	1	440-478
1043	eRF1_1	eRF1 domain 1	0.55	4.9	1	480-502
1043	CHASE3	CHASE3 domain	0.025	9.5	1	489-517
1043	Pox_A_type_inc	1/4 449 468 .. 4 23	0.14	8.3	2	495-517
1043	FH2	Formin Homology 2 Domain	6.2e-154	521.5	1	596-969
1043	DUF387	Putative transcriptional regulators (0.0045	10.2	1	777-800
1043	EMP24_GP25L	emp24/gp25L/p24 family	0.18	5.7	1	868-899
1043	IE68	Herpesvirus immediate early protein	0.25	6.4	1	882-920
1044	KRAB	KRAB box	2.9e-27	100.7	1	8-48
1044	zf-C2H2	Zinc finger, C2H2 type	0.00026	21.9	1	114-136
1044	XPA_N	XPA protein N-terminal	0.51	5.7	1	139-151
1044	TFIIS	Transcription factor S-II (TFIIS)	0.16	7.4	2	142-152
1044	zf-C2H2	Zinc finger, C2H2 type	6.7e-06	28.3	2	142-164
1044	XPA_N	XPA protein N-terminal	0.49	5.8	2	167-179
1044	TFIIS	Transcription factor S-II (TFIIS)	0.18	7.2	3	170-180
1044	zf-C2H2	Zinc finger, C2H2 type	4e-06	29.2	3	170-192
1044	TFIIS	Transcription factor S-II (TFIIS)	0.5	5.7	4	198-208
1044	zf-C2H2	Zinc finger, C2H2 type	6.3e-05	24.4	4	198-220
1044	zf-C2H2	Zinc finger, C2H2 type	7.8e-07	32.1	5	226-248
1044	XPA_N	4/13 223 235 .. 1 13	0.45	5.9	5	251-263
1044	eIF5_eIF2B	Domain found in IF2B/IF5	0.95	3.5	1	254-264
1044	TFIIS	Transcription factor S-II (TFIIS)	0.069	8.7	6	254-264
1044	Transposase_12	Transposase	0.11	5.8	1	254-280
1044	zf-C2H2	Zinc finger, C2H2 type	5.9e-07	32.6	6	254-276
1044	zf-BED	BED zinc finger	0.14	6.9	1	255-277
1044	XPA_N	4/13 223 235 .. 1 13	0.15	7.6	6	279-291
1044	TFIIS	Transcription factor S-II (TFIIS)	0.15	7.5	7	282-292
1044	zf-C2H2	Zinc finger, C2H2 type	8.3e-07	32.0	7	282-304
1044	zf-C2H2	Zinc finger, C2H2 type	4.6e-06	29.0	8	310-332
1044	zf-C2H2	Zinc finger, C2H2 type	5.3e-06	28.7	9	338-360
1044	zf-C2H2	Zinc finger, C2H2 type	4e-06	29.2	10	366-388
1044	XPA_N	4/13 223 235 .. 1 13	0.84	5.0	8	391-403
1044	TFIIS	10/18 368 376 .. 31 39	0.25	6.8	11	394-404
1044	zf-C2H2	Zinc finger, C2H2 type	2.1e-06	30.3	11	394-416
1044	zf-C2H2	Zinc finger, C2H2 type	7.5e-06	28.1	12	422-444
1044	Evr1_Alr	Erv1 / Alr family	0.48	5.4	1	442-460
1044	zf-C2H2	Zinc finger, C2H2 type	9.9e-07	31.7	13	450-472
1044	PqiA	Paraquat-inducible protein A	0.025	8.9	2	469-500

532
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1044	XPA_N	4/13 223 235 .. 1 13	0.45	5.9	9	475-487
1044	eIF5_eIF2B	Domain found in IF2B/IF5	0.95	3.5	2	478-488
1044	TFIIS	12/18 424 432 .. 31 39	0.069	8.7	14	478-488
1044	zf-C2H2	Zinc finger, C2H2 type	2.3e-06	30.2	14	478-500
1044	zf-C2H2	Zinc finger, C2H2 type	6.2e-05	24.4	15	506-528
1044	zf-C2H2	Zinc finger, C2H2 type	1.1e-07	35.5	16	534-556
1044	TFIIS	12/18 424 432 .. 31 39	0.78	5.1	15	536-544
1044	TFIIS	12/18 424 432 .. 31 39	0.12	7.9	16	562-572
1044	zf-C2H2	Zinc finger, C2H2 type	2.5e-06	30.0	17	562-584
1044	TFIIS	12/18 424 432 .. 31 39	0.25	6.8	17	590-600
1044	zf-C2H2	Zinc finger, C2H2 type	6.2e-08	36.5	18	590-612
1044	Umbravirus_LD M	Umbravirus long distance movement (LD)	0.56	2.8	1	601-626
1044	TFIIS	12/18 424 432 .. 31 39	0.062	8.8	18	618-628
1044	zf-C2H2	Zinc finger, C2H2 type	3.6e-07	33.4	19	618-640
1044	zf-BED	3/7 423 445 .. 24 52	0.027	9.3	7	619-641
1045	Sprouty	Sprouty protein (Spry)	1.2e-17	55.0	1	33-70
1045	Sprouty	Sprouty protein (Spry)	2.7e-10	31.5	2	73-90
1046	HAMP	HAMP domain	0.21	7.3	1	9-42
1046	PA	PA domain	3.6e-19	65.4	1	155-255
1046	Peptidase_M28	Peptidase family M28	1.4e-120	410.8	1	332-585
1046	Borrelia_lipo	Borrelia burgdorferi virulent strain	0.98	2.5	1	591-604
1046	TFR_dimer	Transferrin receptor-like dimerisatio	1e-65	228.5	1	597-739
1047	GvpG	Gas vesicle protein G	0.088	6.7	1	17-49
1048	Sema	Sema domain	3.2e-08	29.3	1	34-127
1048	ABM	Antibiotic biosynthesis monooxygenase	0.74	5.7	1	192-208
1048	Sema	Sema domain	2.3e-06	22.6	2	386-449
1048	PSI	Plexin repeat	2.3e-20	65.3	1	468-519
1048	PSI	Plexin repeat	1.4e-12	41.0	2	759-801
1048	TIG	IPT/TIG domain	1.6e-20	78.3	1	803-893
1048	TIG	IPT/TIG domain	4.5e-19	73.5	2	895-980
1048	TIG	IPT/TIG domain	3.3e-13	51.7	3	983-1092
1048	Competence	Competence protein	0.77	3.3	1	1181-1224
1048	RNB	RNB-like protein	0.064	6.4	1	1389-1412
1048	Fimbrial_K88	Fimbrial, major and minor subunit	0.15	5.4	1	1461-1470
1048	ubiquitin	Ubiquitin family	0.021	10.4	1	1463-1497
1049	BTB	BTB/POZ domain	4.5e-28	102.9	1	20-124
1050	ABC_tran	ABC transporter	8.3e-40	134.3	1	26-217
1050	DUF908	Domain of Unknown Function (DUF908)	0.55	3.1	1	69-83
1050	RhoGAP	RhoGAP domain	0.058	7.1	1	69-82
1050	Chlamydia_PMP	Chlamydia polymorphic membrane protei	0.63	2.9	1	546-565
1051	ZZ	Zinc finger, ZZ type	1e-12	48.2	1	3-48
1051	SoxD	Sarcosine oxidase, delta subunit fami	0.97	4.2	1	77-84
1051	zf-C2H2	Zinc finger, C2H2 type	0.00067	20.3	1	78-101
1051	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.3	3.6	1	93-113
1051	SPDY	Domain of unknown function	0.6	4.4	1	117-131

533

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
		(DUF317)				
1051	Di19	Drought induced 19 protein (Di19)	0.00056	13.0	1	312-328
1052	ig	Immunoglobulin domain	7e-12	47.4	1	34-110
1052	MyTH4	MyTH4 domain	0.34	5.6	1	141-152
1052	ig	Immunoglobulin domain	0.0023	15.5	2	150-204
1053	kringle	Kringle domain	0.031	8.7	1	36-64
1053	WSC	WSC domain	2.3e-06	25.6	1	89-142
1053	CUB	CUB domain	1.4e-15	52.5	1	156-260
1054	ig	Immunoglobulin domain	1.7e-05	23.4	1	36-113
1054	Phage_cap_E	Phage major capsid protein E	0.79	2.8	1	127-136
1055	MHC_I	Class I Histocompatibility antigen, d	1e-146	497.6	1	25-203
1055	DUF497	Protein of unknown function (DUF497)	0.2	6.7	1	43-56
1055	ig	Immunoglobulin domain	5.4e-09	36.5	1	220-285
1055	DUF395	YeeE/YedE family (DUF395)	0.18	7.2	1	310-335
1056	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal d	5.8e-05	18.1	1	56-139
1057	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal d	5.1e-38	130.2	1	22-185
1057	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal d	1.6e-12	45.3	1	291-429
1057	Peptidase_M20	Peptidase family M20/M25/M40	0.3	4.0	1	380-419
1058	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal d	5.1e-38	130.2	1	22-185
1058	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal d	5.8e-05	18.1	1	291-374
1059	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal d	3.3e-41	141.1	1	39-201
1059	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal d	5.8e-05	18.1	1	307-390
1060	Secretogranin_V	Neuroendocrine protein 7B2 precursor	2.2e-134	456.6	1	1-204
1060	Ribosomal_L19e	Ribosomal protein L19e	0.7	3.6	1	167-193
1062	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	6.9e-46	159.3	1	4-181
1062	Acyl_transf_3	Acyltransferase family	0.12	6.3	1	106-151
1063	Ribosomal_L29e	Ribosomal L29e protein family	0.0025	12.9	1	21-49
1064	PDZ	PDZ domain (Also known as DHR or GLGF)	7.6e-11	40.0	1	1-84
1064	PDZ	PDZ domain (Also known as DHR or GLGF)	4.2e-10	37.4	2	209-297
1064	PDZ	PDZ domain (Also known as DHR or GLGF)	2.4e-16	59.3	3	310-393
1064	CBM_11	Carbohydrate binding domain (family 1	0.18	5.1	1	360-378
1064	PDZ	PDZ domain (Also known as DHR or GLGF)	7.3e-19	68.1	4	409-490
1064	DUF390	Protein of unknown function (DUF390)	0.82	0.7	1	534-555
1064	PDZ	PDZ domain (Also known as DHR or GLGF)	2.6e-09	34.6	5	694-775
1065	PID	Phosphotyrosine interaction domain (P	3.3e-47	160.5	1	42-168
1066	Galactosyl_T	Galactosyltransferase	0.17	5.8	1	106-116
1066	Chorismate_synt	Chorismate synthase	0.8	1.6	1	291-298
1067	pkinase	Protein kinase domain	1e-73	255.1	1	12-272
1068	lipocalin	Lipocalin / cytosolic fatty-acid bind	7e-38	136.0	1	38-186
1068	Triabin	Triabin	0.0018	12.1	1	119-136
1069	lactamase_B	Metallo-beta-lactamase superfamily	1.7e-21	80.0	1	11-172
1070	annexin	Annexin	9.9e-30	107.8	1	58-124
1070	annexin	Annexin	6.3e-33	119.1	2	130-196
1070	annexin	Annexin	9.7e-28	100.7	3	213-280

534

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1070	annexin	Annexin	4.3e-33	119.8	4	289-355
1071	SNF	Sodium:neurotransmitter symporter fam	0	1200.7	1	44-574
1071	ATP-sulfurylase	ATP-sulfurylase	0.28	3.8	1	198-220
1071	DUF900	Protein of unknown function (DUF900)	0.98	2.8	1	408-420
1072	ig	Immunoglobulin domain	5.7e-06	25.2	1	38-122
1073	Glypican	Glypican	3.9e-292	979.9	1	3-566
1074	zf-C2H2	Zinc finger, C2H2 type	1	7.4	1	16-40
1074	rrm	RNA recognition motif. (a.k.a. RRM, R	9.2e-10	36.8	1	58-123
1074	PAP_assoc	PAP/25A associated domain	1.6e-14	51.8	1	490-549
1074	Isochorismatase	Isochorismatase family	0.49	4.1	1	700-736
1075	PgpA	Phosphatidylglycerophosphatase A	0.92	3.0	1	12-27
1075	SLT	Transglycosylase SLT domain	0.23	6.2	1	82-112
1075	cNMP_binding	Cyclic nucleotide-binding domain	0.67	4.9	1	173-196
1075	Glyco_transf_29	Glycosyltransferase family 29 (sialyl	1.6e-22	77.6	1	289-506
1076	Sec23_trunk	Sec23/Sec24 trunk domain	0.47	4.0	1	42-53
1076	Hydrolase	haloacid dehalogenase-like hydrolase	0.77	3.7	1	46-76
1078	A2M_N	Alpha-2-macroglobulin family N-termin	4.5e-91	312.7	1	6-613
1078	Big_1	Bacterial Ig-like domain (group 1)	0.62	3.9	1	382-403
1078	A2M	Alpha-2-macroglobulin family	6.2e-64	214.2	1	721-949
1078	A2M	Alpha-2-macroglobulin family	6.2e-132	444.2	2	983-1469
1078	Pox_D2	Pox virus D2 protein	0.18	3.4	1	1446-1461
1079	A2M_N	Alpha-2-macroglobulin family N-termin	1.5e-92	317.7	1	19-626
1079	Big_1	Bacterial Ig-like domain (group 1)	0.62	3.9	1	395-416
1079	A2M	Alpha-2-macroglobulin family	3e-44	147.6	1	735-836
1079	kringle	Kringle domain	0.077	7.4	1	840-859
1080	A2M_N	Alpha-2-macroglobulin family N-termin	2e-65	227.5	1	6-548
1080	Big_1	Bacterial Ig-like domain (group 1)	0.62	3.9	1	382-403
1081	A2M_N	Alpha-2-macroglobulin family N-termin	4.5e-91	312.7	1	6-613
1081	Big_1	Bacterial Ig-like domain (group 1)	0.62	3.9	1	382-403
1081	A2M	Alpha-2-macroglobulin family	6.2e-64	214.2	1	721-949
1081	A2M	Alpha-2-macroglobulin family	3.2e-137	462.1	2	983-1469
1081	Pox_D2	Pox virus D2 protein	0.18	3.4	1	1446-1461
1082	A2M_N	Alpha-2-macroglobulin family N-termin	1.5e-92	317.7	1	6-613
1082	Big_1	Bacterial Ig-like domain (group 1)	0.62	3.9	1	382-403
1082	A2M	Alpha-2-macroglobulin family	3e-44	147.6	1	722-823
1082	kringle	Kringle domain	0.077	7.4	1	827-846
1083	COesterase	Carboxylesterase	1.4e-192	649.9	1	8-547
1083	A2M_N	Alpha-2-macroglobulin family N-termin	0.83	2.3	1	12-28
1084	EGF	EGF-like domain	2.8e-05	21.8	1	192-219
1084	laminin_EGF	Laminin EGF-like (Domains III and V)	0.37	5.9	1	208-220
1084	Tautomerase	Tautomerase enzyme	0.14	5.5	1	292-318

535

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1084	EGF	EGF-like domain	9.9e-07	27.0	2	404-431
1084	Arthro_defensin	1/3 199 211 .. 23 36	0.52	3.0	2	411-423
1084	TB	TB domain	1.3e-16	54.5	1	567-610
1084	EGF	EGF-like domain	1.1e-07	30.4	3	631-666
1084	MoeZ_MoeB	MoeZ/MoeB domain	0.62	3.6	1	676-682
1084	TB	TB domain	2.9e-26	85.2	2	688-729
1084	EGF	EGF-like domain	1.9e-07	29.6	4	878-914
1084	TIL	Trypsin Inhibitor like cysteine rich	0.00021	14.8	2	898-920
1084	CBM_14	Chitin binding Peritrophin-A domain	0.88	3.8	1	901-920
1084	EGF	EGF-like domain	1.2e-07	30.3	5	920-956
1084	TIL	Trypsin Inhibitor like cysteine rich	0.00057	13.5	3	941-962
1084	squash	Squash family serine protease inhibit	0.069	4.9	1	942-969
1084	granulin	Granulin	0.06	7.4	1	943-958
1084	EGF	EGF-like domain	0.098	9.0	6	962-983
1084	VSP	Giardia variant-specific surface prot	0.031	7.4	1	982-1003
1084	EGF	EGF-like domain	5.2e-06	24.4	7	1043-1078
1084	EGF	EGF-like domain	1.6e-06	26.2	8	1084-1119
1084	TIL	5/14 1063 1084 .. 47 68	0.98	3.3	6	1103-1125
1084	VSP	Giardia variant-specific surface prot	0.29	3.9	2	1105-1126
1084	EGF	EGF-like domain	2.1e-06	25.9	9	1125-1160
1084	TIL	5/14 1063 1084 .. 47 68	0.0095	9.6	7	1145-1166
1084	EGF	EGF-like domain	5.9e-05	20.6	10	1166-1201
1084	Plasmod_Pvs28	Plasmodium ookinete surface protein P	0.11	6.2	1	1172-1210
1084	TIL	5/14 1063 1084 .. 47 68	0.0044	10.7	8	1187-1207
1084	EGF	EGF-like domain	5.5e-06	24.3	11	1207-1243
1084	PAD_porph	Porphyromonas-type peptidyl-arginine	0.047	8.4	1	1224-1234
1084	Plasmod_Pvs28	Plasmodium ookinete surface protein P	0.57	3.6	2	1226-1281
1084	TIL	5/14 1063 1084 .. 47 68	0.016	8.9	9	1228-1249
1084	VSP	Giardia variant-specific surface prot	0.043	6.9	3	1229-1249
1084	EGF	EGF-like domain	4.9e-06	24.5	12	1249-1285
1084	EGF	EGF-like domain	1.1e-05	23.3	13	1291-1328
1084	TB	TB domain	8.7e-23	74.2	3	1358-1401
1084	EGF	EGF-like domain	6e-05	20.6	14	1429-1466
1084	CBM_14	Chitin binding Peritrophin-A domain	0.3	5.3	2	1452-1472
1084	EB	2/4 962 979 .. 1 18	0.79	4.0	3	1472-

536
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						1483
1084	EGF	EGF-like domain	1.3e-07	30.3	15	1472-1507
1084	TB	TB domain	1.8e-23	76.3	4	1535-1577
1084	EGF	EGF-like domain	3.3e-06	25.1	16	1626-1661
1084	Plasmod_Pvs28	Plasmodium ookinete surface protein P	0.44	4.0	3	1632-1708
1084	TIL	5/14 1063 1084 .. 47 68	0.00068	13.2	13	1642-1667
1084	VSP	Giardia variant-specific surface prot	0.96	2.0	4	1646-1667
1084	EGF	EGF-like domain	1.2e-07	30.3	17	1667-1706
1084	TIL	5/14 1063 1084 .. 47 68	0.85	3.5	14	1690-1706
1086	ig	Immunoglobulin domain	5.3e-07	29.1	1	168-232
1086	Corona_NS4	Coronavirus non-structural protein NS	0.47	3.5	1	248-271
1086	ig	Immunoglobulin domain	0.052	10.4	2	285-347
1086	fn3	Fibronectin type III domain	2.4e-16	58.5	1	373-459
1086	fn3	Fibronectin type III domain	3.1e-15	54.7	2	501-587
1086	fn3	Fibronectin type III domain	5.5e-19	67.7	3	602-685
1086	fn3	Fibronectin type III domain	1e-12	45.9	4	700-786
1086	fn3	Fibronectin type III domain	1.7e-27	97.2	5	802-888
1086	OsmC	OsmC-like protein	0.57	4.1	1	984-1018
1086	ig	Immunoglobulin domain	0.00041	18.2	3	1133-1191
1086	ig	Immunoglobulin domain	1.7e-07	30.9	4	1349-1405
1086	Aegerolysin	Aegerolysin	0.56	4.3	1	1411-1428
1087	KRAB	KRAB box	5.8e-25	92.4	1	14-54
1087	DUF19	Domain of unknown function DUF19	0.044	5.2	1	80-105
1087	TFIIS	Transcription factor S-II (TFIIS)	1	4.7	1	161-171
1087	zf-C2H2	Zinc finger, C2H2 type	8.3e-07	32.0	1	161-183
1087	XPA_N	XPA protein N-terminal	0.49	5.8	2	186-198
1087	TFIIS	Transcription factor S-II (TFIIS)	0.21	7.0	2	189-199
1087	zf-C2H2	Zinc finger, C2H2 type	2.8e-07	33.9	2	189-211
1087	XPA_N	XPA protein N-terminal	0.47	5.9	3	214-226
1087	zf-C2H2	Zinc finger, C2H2 type	5.4e-07	32.7	3	217-239
1087	zf-BED	BED zinc finger	0.11	7.3	2	222-240
1087	XPA_N	XPA protein N-terminal	0.52	5.7	4	242-254
1087	TFIIS	Transcription factor S-II (TFIIS)	0.32	6.4	4	245-255
1087	zf-C2H2	Zinc finger, C2H2 type	3.2e-08	37.5	4	245-267
1087	zf-BED	BED zinc finger	0.16	6.8	3	246-268
1088	KRAB	KRAB box	5.8e-25	92.4	1	14-54
1088	DUF19	Domain of unknown function DUF19	0.044	5.2	1	80-105
1088	TFIIS	Transcription factor S-II (TFIIS)	1	4.7	1	161-171
1088	zf-C2H2	Zinc finger, C2H2 type	4.1e-08	37.1	1	161-183
1088	zf-BED	BED zinc finger	0.24	6.2	1	162-184
1089	Keratin_B2	Keratin, high sulfur B2 protein	5.6e-21	73.7	1	2-117
1089	Keratin_B2	Keratin, high sulfur B2 protein	0.011	9.4	2	118-170

537

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1090	Keratin_B2	Keratin, high sulfur B2 protein	5.9e-11	38.5	1	2-75
1091	Keratin_B2	Keratin, high sulfur B2 protein	7.5e-06	20.6	1	2-40
1091	Keratin_B2	Keratin, high sulfur B2 protein	4e-18	63.7	2	41-144
1091	Keratin_B2	Keratin, high sulfur B2 protein	4e-05	18.0	3	145-205
1092	abhydro_lipase	ab-hydrolase associated lipase region	1.9e-32	117.8	1	27-97
1092	abhydrolase	alpha/beta hydrolase fold	9.5e-19	67.6	1	111-388
1093	abhydro_lipase	ab-hydrolase associated lipase region	1.9e-32	117.8	1	87-157
1093	abhydrolase	alpha/beta hydrolase fold	9.5e-19	67.6	1	171-448
1094	7tm_3	7 transmembrane receptor (metabotropi	0.75	3.2	1	24-45
1094	7tm_3	7 transmembrane receptor (metabotropi	0.00057	14.4	2	65-109
1094	Condensation	Condensation domain	0.36	4.2	1	157-169
1094	7tm_3	7 transmembrane receptor (metabotropi	2.1e-05	19.4	3	168-271
1095	Tuberin	Tuberin	0.59	0.5	1	17-23
1095	DAGAT	Diacylglycerol acyltransferase	6.2e-98	335.5	1	38-216
1096	GASA	Gibberellin regulated protein	0.35	1.3	1	22-51
1096	lectin_c	Lectin C-type domain	8.9e-26	95.8	1	100-208
1097	GASA	Gibberellin regulated protein	0.35	1.3	1	22-51
1097	lectin_c	Lectin C-type domain	2.5e-27	101.0	1	100-208
1098	7tm_1	7 transmembrane receptor (rhodopsin f	2.6e-50	149.5	1	41-290
1098	endotoxin_N	delta endotoxin, N-terminal domain	0.87	3.6	1	195-225
1099	SEA	SEA domain	1.2e-06	24.2	1	330-408
1099	DUF916	Bacterial protein of unknown function	0.091	7.2	1	550-576
1099	Hanta_G2	Hantavirus glycoprotein G2	0.027	6.9	1	550-578
1099	Peptidase_C13	Peptidase C13 family	0.32	3.5	1	554-582
1100	ig	Immunoglobulin domain	0.0046	14.3	1	146-203
1100	ig	Immunoglobulin domain	3.2e-07	29.9	2	245-295
1100	FHIPEP	FHIPEP family	0.21	3.5	1	315-326
1101	LRRNT	Leucine rich repeat N-terminal domain	0.00068	15.2	1	23-49
1101	LRR	Leucine Rich Repeat	8.7e-05	18.9	1	51-74
1101	LRR	Leucine Rich Repeat	0.00032	17.0	2	75-98
1101	LRR	Leucine Rich Repeat	0.025	10.6	3	99-122
1101	LRR	Leucine Rich Repeat	0.00069	15.8	4	123-146
1101	LRR	Leucine Rich Repeat	9.9e-06	22.1	5	147-170
1101	LRRCT	Leucine rich repeat C-terminal domain	2.3e-15	48.2	1	180-232
1101	ig	Immunoglobulin domain	1.3e-08	35.1	1	248-307
1101	ig	Immunoglobulin domain	3.8e-09	37.1	2	344-400
1101	ig	Immunoglobulin domain	3.4e-05	22.3	3	440-490
1101	BON	Transport-associated domain	0.14	7.1	1	495-507
1101	ig	Immunoglobulin domain	3.1e-08	33.7	4	525-582
1101	pec_lyase_N	Pectate lyase, N terminus	0.19	3.9	1	626-632
1101	An_peroxidase	Animal haem peroxidase	9.8e-195	657.1	1	726-1265
1101	PAL	Phenylalanine and histidine ammonia-l	0.53	2.6	1	993-1010
1101	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	1057-1065
1101	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	1150-1167
1101	TILa	TILa domain	0.00018	16.9	1	1394-1433
1101	PSP94	Beta-microseminoprotein (PSP-94)	0.11	8.0	1	1395-1426
1101	vwc	von Willebrand factor type C domain	2e-10	38.0	1	1395-1450

538

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1102	LRRNT	Leucine rich repeat N-terminal domain	0.00068	15.2	1	23-49
1102	LRR	Leucine Rich Repeat	8.7e-05	18.9	1	51-74
1102	LRR	Leucine Rich Repeat	0.021	10.9	2	75-98
1102	LRR	Leucine Rich Repeat	0.00069	15.8	3	99-122
1102	LRR	Leucine Rich Repeat	9.9e-06	22.1	4	123-146
1102	LRRCT	Leucine rich repeat C-terminal domain	2.3e-15	48.2	1	156-208
1102	ig	Immunoglobulin domain	1.3e-08	35.1	1	224-283
1102	ig	Immunoglobulin domain	3.8e-09	37.1	2	320-376
1102	ig	Immunoglobulin domain	3.4e-05	22.3	3	416-466
1102	BON	Transport-associated domain	0.14	7.1	1	471-483
1102	ig	Immunoglobulin domain	3.1e-08	33.7	4	501-558
1102	pec_lyase_N	Pectate lyase, N terminus	0.19	3.9	1	602-608
1102	An_peroxidase	Animal haem peroxidase	9.8e-195	657.1	1	702-1241
1102	PAL	Phenylalanine and histidine ammonia-l	0.53	2.6	1	969-986
1102	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	1033-1041
1102	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	1126-1143
1102	TILa	TILa domain	0.00018	16.9	1	1370-1409
1102	PSP94	Beta-microseminoprotein (PSP-94)	0.11	8.0	1	1371-1402
1102	vwc	von Willebrand factor type C domain	2e-10	38.0	1	1371-1426
1103	PMEI	Plant invertase/pectin methylesterase	0.33	5.5	1	2-24
1103	Ribosomal_S26e	Ribosomal protein S26e	0.47	3.9	1	215-236
1103	ATP-gua_Ptrans	ATP:guanido phosphotransferase, C-ter	0.089	6.1	1	240-262
1103	Arch fla_DE	Archaeal flagella protein	0.42	5.0	1	670-683
1103	zf_dskA_traR	Prokaryotic dksA/traR C4-type zinc fi	0.48	4.8	1	1145-1160
1103	zf-C4	Zinc finger, C4 type (two domains)	0.07	7.6	1	1147-1157
1104	UBX	UBX domain	0.79	4.8	1	1-18
1104	FTCD_C	Formiminotransferase-cyclodeaminase	0.21	6.0	1	47-77
1104	IKI3	IKI3 family	0.66	0.9	1	80-94
1104	Ribosomal_L21p	Ribosomal prokaryotic L21 protein	0.27	5.0	1	115-138
1104	DUF709	Family of unknown function (DUF709)	0.28	6.5	1	215-227
1105	Torsin	Torsin	8.2e-07	23.3	1	106-125
1106	rrm	RNA recognition motif. (a.k.a. RRM, R	0.086	8.5	1	41-71
1107	MH1	MH1 domain	0.16	5.4	1	288-309
1107	Pecanex_C	Pecanex protein (C-terminus)	2.2e-129	440.0	1	437-621
1107	Pecanex_C	Pecanex protein (C-terminus)	4e-08	29.6	2	622-640
1108	MH1	MH1 domain	0.16	5.4	1	288-309
1108	Pecanex_C	Pecanex protein (C-terminus)	1.2e-106	364.5	1	437-599
1110	Herpes_UL14	Herpesvirus UL14-like protein	0.12	6.1	1	17-43
1110	ig	Immunoglobulin domain	0.57	6.5	1	36-55
1110	Ribosomal_L4	Ribosomal protein L4/L1 family	0.95	3.3	1	79-107
1110	Vpu	Vpu protein	0.34	4.7	1	106-140
1110	BPD_transp	Binding-protein-dependent transport s	0.9	3.7	1	109-137
1111	DUF895	Eukaryotic protein of unknown functio	0.68	4.1	1	133-149
1112	RWD	RWD domain	9.8e-40	142.2	1	11-125

539

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1112	globin	Globin	0.048	8.6	1	88-120
1112	eRF1_2	eRF1 domain 2	0.72	4.1	1	114-127
1112	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	7.5e-10	27.9	1	135-201
1112	DNA_ligase_ZB D	NAD-dependent DNA ligase C4 zinc fing	0.37	5.8	1	196-207
1112	zf-MIZ	MIZ zinc finger	0.28	4.6	1	197-207
1112	ApoA-II	Apolipoprotein A-II (ApoA-II)	0.94	3.6	1	261-272
1113	pkinase	Protein kinase domain	1e-45	162.0	1	194-468
1113	Pox_M2	Poxvirus M2 protein	0.82	3.3	1	306-333
1113	DUF857	Domain of unknown function (DUF857)	0.48	4.7	1	417-428
1113	ExoD	Exopolysaccharide synthesis, ExoD	0.65	2.7	1	424-456
1113	KE2	KE2 family protein	0.18	7.7	1	468-502
1114	SNF2_N	SNF2 family N-terminal domain	0.057	6.4	1	82-115
1114	Transposase_8	Transposase	0.38	5.6	1	82-105
1114	OKR_DC_1_N	Orn/Lys/Arg decarboxylase, N-terminal	0.34	3.6	1	91-102
1116	DUF1006	Protein of unknown function (DUF1006)	0.45	2.5	1	5-24
1117	ig	Immunoglobulin domain	2.3e-05	22.9	1	30-87
1117	ig	Immunoglobulin domain	0.0023	15.5	2	127-186
1117	ig	Immunoglobulin domain	0.00079	17.2	3	281-337
1117	ig	Immunoglobulin domain	0.026	11.5	4	379-434
1117	SNF7	SNF7	0.95	3.5	1	435-450
1118	ig	Immunoglobulin domain	0.00079	17.2	1	42-98
1118	ig	Immunoglobulin domain	0.026	11.5	2	140-195
1118	SNF7	SNF7	0.95	3.5	1	196-211
1119	IBN_NT	Importin-beta N-terminal domain	1.6e-24	89.3	1	28-100
1119	PurA	PurA ssDNA and RNA-binding protein	0.19	4.8	1	155-171
1119	PAN	PAN domain	1	3.2	1	706-735
1120	Bowman-Birk_leg	Bowman-Birk serine protease inhibitor	1	4.0	1	28-36
1120	RNA_pol_Rpb2_1	RNA polymerase beta subunit	0.25	2.1	1	150-946
1120	cobW	Cobalamin synthesis protein/P47K	0.85	2.3	1	170-205
1120	DUF909	Bacterial protein of unknown function	0.16	7.1	1	215-247
1120	Glyco_hydro_2_C	Glycosyl hydrolases family 2, TIM bar	0.24	4.8	1	262-277
1120	ank	Ankyrin repeat	1.2e-10	41.3	1	920-952
1120	ank	Ankyrin repeat	2.5e-08	33.0	2	953-985
1120	SH3	SH3 domain	5.7e-16	61.3	1	1022-1079
1122	TPR	TPR Domain	0.013	12.3	1	138-157
1122	TPR	TPR Domain	1.1e-07	30.1	2	158-191
1122	TPR	TPR Domain	0.29	7.5	3	192-222
1122	BEX	Brain expressed X-linked like family	0.25	3.9	1	261-294
1122	eRF1_2	eRF1 domain 2	0.12	6.9	1	322-338
1122	Subtilisin_N	Subtilisin N-terminal Region	0.83	5.1	1	323-344
1123	Pencillinase_R	Penicillinase repressor	0.85	4.2	1	57-75
1124	ank	Ankyrin repeat	1.8e-07	29.9	1	64-96
1124	ank	Ankyrin repeat	1.5e-06	26.5	2	97-129
1124	ank	Ankyrin repeat	2e-07	29.7	3	130-162
1124	Shigella_OspC	Shigella flexneri OspC protein	0.51	3.2	1	131-161
1124	ank	Ankyrin repeat	4.3e-06	24.9	4	163-195

540

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1124	ank	Ankyrin repeat	0.00018	19.2	5	196-228
1125	ank	Ankyrin repeat	1.8e-07	29.9	1	64-96
1125	ank	Ankyrin repeat	1.5e-06	26.5	2	97-129
1125	ank	Ankyrin repeat	2e-07	29.7	3	130-162
1125	Shigella_OspC	Shigella flexneri OspC protein	0.51	3.2	1	131-161
1125	ank	Ankyrin repeat	4.3e-06	24.9	4	163-195
1126	DUF846	Eukaryotic protein of unknown function	0.5	2.6	1	50-74
1129	Apolipoprotein	Apolipoprotein A1/A4/E family	0.95	3.4	1	4-28
1129	F5_F8_type_C	F5/8 type C domain	1e-63	195.2	1	34-174
1129	laminin_G	Laminin G domain	2.7e-10	36.3	1	212-344
1130	Apolipoprotein	Apolipoprotein A1/A4/E family	0.95	3.4	1	4-28
1130	F5_F8_type_C	F5/8 type C domain	1e-63	195.2	1	34-174
1130	laminin_G	Laminin G domain	6.5e-11	38.5	1	212-344
1130	laminin_G	Laminin G domain	1.8e-11	40.4	2	398-525
1130	EGF	EGF-like domain	1.1e-06	26.8	1	551-583
1130	fibrinogen_C	Fibrinogen beta and gamma chains, C-t	0.051	6.6	1	601-634
1130	laminin_G	Laminin G domain	2.7e-17	60.5	3	821-943
1130	EGF	EGF-like domain	0.0014	15.7	2	962-996
1130	laminin_G	Laminin G domain	0.00033	15.3	4	1046-1179
1130	DNA_PPF	DNA polymerase processivity factor	0.69	4.3	1	1059-1078
1130	BenE	Benzoate membrane transport protein	0.29	3.8	1	1239-1255
1130	BPD_transp	Binding-protein-dependent transport s	0.41	5.0	1	1245-1276
1131	HTH_9	N-terminal HTH domain of molybdenum-b	0.72	4.3	1	61-84
1131	Glycos_transf_2	Glycosyl transferase	2.2e-31	105.9	1	155-341
1131	Ribosomal_S3_C	Ribosomal protein S3, C-terminal domain	0.98	3.4	1	357-363
1131	Ricin_B_lectin	QXW lectin repeat	0.13	8.4	1	467-496
1131	Ricin_B_lectin	QXW lectin repeat	0.00073	16.5	2	558-596
1132	Enterotoxin_HS	Heat-stable enterotoxin	0.71	1.4	1	37-43
1132	VSP	Giardia variant-specific surface protein	0.23	4.3	1	96-128
1132	tsp_1	Thrombospondin type 1 domain	0.27	6.0	1	148-203
1133	pkinase	Protein kinase domain	1.3e-48	171.6	1	11-237
1133	Pox_ser-thr_kin	Poxvirus serine/threonine protein kinase	0.2	4.5	1	133-156
1133	pkinase	Protein kinase domain	0.00017	16.2	2	322-347
1135	C2	C2 domain	2.4e-30	103.9	1	7-88
1135	Transposase_24	Plant transposase (Pta/En/Spm family)	0.22	4.9	1	42-56
1135	photoRC	Photosynthetic reaction centre protein	0.95	1.8	1	45-67
1135	C2	C2 domain	8e-32	108.9	2	135-216
1135	RasGAP	GTPase-activator protein for Ras-like	5.4e-39	124.6	1	323-513
1135	AraC_binding	AraC-like ligand binding domain	0.42	5.2	1	414-452
1135	PH	PH domain	6.8e-11	37.2	1	567-673
1135	BTK	BTK motif	1.9e-06	26.9	1	675-711
1137	toxin_2	Scorpion short toxin	0.089	6.2	1	51-77
1137	C_tripleX	Cysteine rich repeat	9e-05	15.9	1	54-71
1137	EGF	EGF-like domain	0.00049	17.3	1	60-86
1137	laminin_EGF	Laminin EGF-like (Domains III and V)	0.55	5.3	1	75-88
1137	EGF	EGF-like domain	0.00015	19.2	2	123-155
1137	TIL	Trypsin Inhibitor like cysteine rich	0.55	4.1	2	142-163
1137	EGF	EGF-like domain	0.00018	18.9	3	163-197

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1137	TIL	Trypsin Inhibitor like cysteine rich	0.0065	10.1	3	180-203
1137	EGF	EGF-like domain	0.031	10.8	4	215-232
1137	EGF	EGF-like domain	3.7e-07	28.6	5	248-283
1137	EB	EB module	0.73	4.1	1	254-283
1137	PRK	Phosphoribulokinase / Uridine kinase	0.74	2.8	1	407-426
1137	MAM	MAM domain	3.1e-27	100.7	1	452-593
1137	Omptin	Omptin family	0.99	2.0	1	460-476
1138	SurE	Survival protein SurE	0.68	2.6	1	10-23
1138	Pox_A11	Poxvirus A11 Protein	0.17	3.2	1	57-75
1138	zf-C2H2	Zinc finger, C2H2 type	0.55	8.5	1	205-228
1138	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.64	2.6	1	205-210
1139	4HBT	Thioesterase superfamily	0.038	8.2	1	52-120
1143	7tm_1	7 transmembrane receptor (rhodopsin f	1.4e-28	84.3	1	1-173
1144	7tm_1	7 transmembrane receptor (rhodopsin f	4.5e-49	145.7	1	40-287
1146	SNARE	SNARE domain	0.28	7.0	1	195-229
1147	IL1	Interleukin-1 / 18	2.6e-23	83.4	1	51-152
1148	filament	Intermediate filament protein	3.2e-106	363.0	1	1-296
1148	K-box	K-box region	0.78	4.1	1	14-31
1148	Ribosomal_S4	Ribosomal protein S4/S9 N-terminal do	0.85	4.3	1	76-97
1148	IATP	Mitochondrial ATPase inhibitor, IATP	0.46	6.1	1	125-148
1148	ERG2_Sigma1R	ERG2 and Sigma1 receptor like protein	0.36	3.7	1	162-191
1148	filament	Intermediate filament protein	4.8e-31	113.3	2	380-469
1148	K-box	K-box region	0.11	7.0	2	397-415
1148	bZIP	1/2 51 88 .. 28 65	0.3	6.3	2	435-472
1148	Tfb2	Transcription factor Tfb2	0.17	0.5	1	466-472
1149	Phage_X	Phage X family	0.71	4.2	1	16-41
1149	2OG-FeII_Oxy	2OG-Fe(II) oxygenase superfamily	0.27	6.0	1	229-307
1150	MBOAT	MBOAT family	2.3e-08	30.9	1	90-249
1151	filament	Intermediate filament protein	1.2e-38	138.5	1	131-242
1151	filament	Intermediate filament protein	2.9e-83	286.8	2	244-412
1151	HSP70	Hsp70 protein	0.99	2.0	1	268-294
1151	HAMP	HAMP domain	1	4.8	1	301-334
1151	DUF164	Uncharacterized ACR, COG1579	0.057	7.3	1	310-352
1151	bZIP	bZIP transcription factor	0.062	8.7	2	316-348
1151	Transposase_8	Transposase	0.79	4.5	1	317-335
1151	MutS_V	MutS domain V	0.27	4.5	1	354-370
1151	OEP	Outer membrane efflux protein	0.053	7.0	1	356-393
1151	MutS_IV	MutS family domain IV	0.9	4.6	1	359-392
1151	Hpt	Hpt domain	0.49	5.2	1	365-389
1151	Retro_M	Retroviral M domain	0.5	4.2	1	369-377
1152	Peptidase_M10_N	Matrix metalloprotease, N-terminal do	1e-42	119.7	1	12-95
1152	PG_binding_1	Putative peptidoglycan binding domain	0.022	10.3	1	60-90
1152	Peptidase_M10	Matrixin	8.7e-51	178.9	1	102-206
1152	hemopexin	Hemopexin	1.6e-08	30.8	1	231-273
1152	hemopexin	Hemopexin	4.4e-11	39.4	2	275-317
1152	hemopexin	Hemopexin	2.2e-13	47.1	3	322-367
1152	hemopexin	Hemopexin	2e-05	20.4	4	371-411
1153	Peptidase_M10_N	Matrix metalloprotease, N-terminal do	1e-42	119.7	1	12-95
1153	PG_binding_1	Putative peptidoglycan binding domain	0.022	10.3	1	60-90
1153	Peptidase_M10	Matrixin	8.7e-51	178.9	1	102-206
1153	hemopexin	Hemopexin	1.6e-08	30.8	1	231-273

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1153	hemopexin	Hemopexin	4.4e-11	39.4	2	275-317
1153	hemopexin	Hemopexin	2.2e-13	47.1	3	322-367
1153	hemopexin	Hemopexin	2e-05	20.4	4	371-411
1154	SUFU	Suppressor of fused protein (SUFU)	0	1218.3	1	3-484
1155	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal d	1.5e-61	210.5	1	38-217
1155	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal d	4.6e-32	115.6	1	242-478
1156	HMG_box	HMG (high mobility group) box	5.8e-32	115.2	1	85-153
1156	HEV_ORF2	Hepatitis E virus ORF-2 (Putative cap	0.026	8.5	1	336-356
1159	zf-PARP	Poly(ADP-ribose) polymerase and DNA-L	3.5e-52	183.5	1	93-185
1159	DNA_ligase_A_N	DNA ligase N terminus	5.9e-13	42.4	1	319-433
1159	DNA_ligase	ATP dependent DNA ligase domain	8.5e-74	255.3	1	480-636
1159	mRNA_cap_enzyme	mRNA capping enzyme, catalytic domain	0.00064	7.5	1	594-613
1160	serpin	Serpin (serine protease inhibitor)	9.5e-151	511.0	1	1-425
1166	Peptidase_C14	Caspase domain	2.4e-06	23.6	1	7-40
1167	ig	Immunoglobulin domain	2e-05	23.2	1	42-96
1167	ig	Immunoglobulin domain	0.0012	16.6	2	135-197
1167	ig	Immunoglobulin domain	0.0013	16.4	3	237-297
1168	UK	Virulence determinant	0.083	7.0	1	14-38
1168	TIP120	TBP (TATA-binding protein) -interacti	0	2347.3	1	25-908
1168	HEAT	HEAT repeat	0.093	8.3	2	248-286
1168	HEAT	HEAT repeat	0.022	10.4	3	343-364
1168	Armadillo_seg	Armadillo/beta-catenin-like repeat	0.2	8.0	2	682-721
1169	lectin_c	Lectin C-type domain	7.4e-19	72.8	1	131-231
1171	WD40	WD domain, G-beta repeat	8.5e-11	40.4	1	223-260
1171	WD40	WD domain, G-beta repeat	2.8e-06	24.7	2	280-316
1171	WD40	WD domain, G-beta repeat	9e-09	33.4	3	320-357
1171	WD40	WD domain, G-beta repeat	0.0041	13.7	4	362-398
1171	WD40	WD domain, G-beta repeat	3.1e-14	52.4	5	403-440
1171	WD40	WD domain, G-beta repeat	4.3e-12	45.0	6	445-491
1171	WD40	WD domain, G-beta repeat	1.1e-14	54.0	7	496-533
1171	WD40	WD domain, G-beta repeat	0.23	7.6	8	538-574
1172	ig	Immunoglobulin domain	2.3e-05	22.9	1	42-99
1172	ig	Immunoglobulin domain	0.0023	15.5	2	139-198
1172	MBOAT	MBOAT family	3.1e-07	26.8	1	741-769
1173	ig	Immunoglobulin domain	2.3e-05	22.9	1	42-99
1173	ig	Immunoglobulin domain	0.0023	15.5	2	139-198
1173	MBOAT	MBOAT family	3.1e-07	26.8	1	741-769
1174	MBOAT	MBOAT family	2.3e-08	30.9	1	90-249
1174	MBOAT	MBOAT family	4.7e-09	33.5	2	308-351
1175	PTE	Phosphotriesterase family	1.2e-139	474.0	1	7-233
1182	Pox_int_trans	Poxvirus intermediate transcription fac	0.092	5.7	1	94-122
1183	PS_Dcarboxylase	Phosphatidylserine decarboxylase	0.011	11.4	1	165-181
1183	PS_Dcarboxylase	Phosphatidylserine decarboxylase	8.3e-52	182.3	2	246-467
1184	TSC22	TSC-22/dip/bun family	5e-47	146.4	1	124-183
1186	ADP_PFK_GK	ADP-specific Phosphofructokinase/Glucokin	7e-227	763.9	1	68-492

543

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1186	Mannitol_dh	Mannitol dehydrogenase	0.052	6.8	1	394-413
1187	DENN	DENN (AEX-3) domain	0.054	6.0	1	16-40
1188	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	0.5	1.1	1	310-346
1188	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	2.4e-07	19.7	2	513-608
1188	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	5.3e-08	21.7	3	646-680
1188	Peptidase_S9	Prolyl oligopeptidase family	3.9e-11	36.8	1	692-764
1188	Esterase	Putative esterase	0.062	6.6	1	738-781
1189	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	0.5	1.1	1	310-346
1189	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	2.4e-07	19.7	2	513-608
1189	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	5.3e-08	21.7	3	646-680
1189	Peptidase_S9	Prolyl oligopeptidase family	3.9e-11	36.8	1	692-764
1189	Esterase	Putative esterase	0.062	6.6	1	738-781
1190	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	0.5	1.1	1	310-346
1190	DPPIV_N_term	Dipeptidyl peptidase IV (DPP IV) N-termi	5.3e-08	21.7	2	633-667
1190	Peptidase_S9	Prolyl oligopeptidase family	3.9e-11	36.8	1	679-751
1190	Esterase	Putative esterase	0.062	6.6	1	725-768
1191	Ribosomal_S25	S25 ribosomal protein	6.7e-67	232.4	1	2-100
1191	DUF387	Putative transcriptional regulators (Yp	0.099	5.8	1	65-87
1193	ank	Ankyrin repeat	6.7e-10	38.6	2	49-81
1193	ank	Ankyrin repeat	2.7e-08	32.8	3	82-114
1193	ank	Ankyrin repeat	0.0036	14.4	4	115-147
1193	ank	Ankyrin repeat	1.5e-11	44.5	5	148-180
1193	ank	Ankyrin repeat	1.2e-08	34.1	6	181-213
1193	ank	Ankyrin repeat	3.3e-08	32.5	7	214-246
1193	ank	Ankyrin repeat	3.4e-11	43.2	8	247-279
1193	ank	Ankyrin repeat	1.3e-08	33.9	9	280-313
1193	ank	Ankyrin repeat	0.0027	14.9	10	314-346
1193	ank	Ankyrin repeat	8.5e-08	31.1	11	347-379
1193	ank	Ankyrin repeat	0.013	12.4	12	380-404
1193	ank	Ankyrin repeat	8.3e-08	31.1	13	431-463
1193	ank	Ankyrin repeat	1.1e-09	37.8	14	464-496
1193	ank	Ankyrin repeat	6.9e-07	27.8	15	497-557
1193	endonuclease_7	Recombination endonuclease VII	0.034	9.6	1	513-537
1193	ank	Ankyrin repeat	0.0047	14.0	16	558-581
1193	ank	Ankyrin repeat	1.2e-05	23.3	17	596-625
1193	ank	Ankyrin repeat	3.8e-10	39.4	18	626-658
1193	ank	Ankyrin repeat	0.00034	18.1	19	660-692
1193	ank	Ankyrin repeat	1.5e-09	37.3	20	696-728
1193	ank	Ankyrin repeat	1.5e-05	23.0	21	729-761
1193	ank	Ankyrin repeat	9.7e-05	20.1	22	762-784
1193	ank	Ankyrin repeat	2.3e-07	29.5	23	798-821
1193	ank	Ankyrin repeat	0.0023	15.1	24	830-853
1193	ank	Ankyrin repeat	3.6e-10	39.5	25	865-897
1193	ank	Ankyrin repeat	4.3e-06	24.9	26	898-931
1193	ank	Ankyrin repeat	0.00019	19.1	27	932-964
1193	ank	Ankyrin repeat	1.6e-07	30.0	28	968-

544
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						1000
1194	trypsin	Trypsin	1.5e-23	72.8	1	166-342
1194	PDZ	PDZ domain (Also known as DHR or GLGF)	0.0018	14.0	1	372-412
1195	7tm_1	7 transmembrane receptor (rhodopsin family)	2.4e-18	53.6	1	1-137
1196	vwc	von Willebrand factor type C domain	5.4e-05	19.0	1	66-105
1196	vwc	von Willebrand factor type C domain	0.16	6.8	2	108-163
1196	vwc	von Willebrand factor type C domain	2e-07	27.4	3	166-192
1197	7tm_1	7 transmembrane receptor (rhodopsin family)	4.5e-36	106.7	1	46-295
1198	MethyltransfD12	D12 class N6 adenine-specific DNA met	2.1e-36	127.8	1	30-152
1199	lipocalin	Lipocalin / cytosolic fatty-acid binding pr	1.8e-21	80.6	1	32-176
1200	tRNA_anti	OB-fold nucleic acid binding domain	2.7e-15	56.4	1	44-118
1200	tRNA-synt_2	tRNA synthetases class II (D, K and N)	2.7e-91	313.5	1	135-473
1202	FAD_binding_2	FAD binding domain	5.9e-54	182.8	1	5-100
1203	RasGEFN	Guanine nucleotide exchange factor for Ras-l	0.00072	15.7	1	39-87
1203	RasGEF	RasGEF domain	0.39	5.5	1	211-240
1203	RasGEF	RasGEF domain	6.8e-18	69.6	2	280-360
1204	KH	KH domain	3.8e-17	61.6	1	17-63
1204	KH	KH domain	5.4e-19	68.0	2	101-150
1204	KH	KH domain	5.8e-16	57.6	3	265-313
1206	transket_pyr	Transketolase, pyridine binding domai	3.3e-75	258.0	1	15-190
1206	transketolase_C	Transketolase, C-terminal domain	2e-59	194.9	1	208-331
1207	Calsequestrin	Calsequestrin	6.3e-294	986.6	1	5-390
1207	thioered	Thioredoxin	0.057	9.0	1	123-152
1209	PH	PH domain	0.0057	11.0	1	70-97
1210	ig	Immunoglobulin domain	4.9e-07	29.2	1	35-112
1210	ig	Immunoglobulin domain	2e-06	26.9	2	154-228
1213	cadherin	Cadherin domain	0.00026	16.8	1	33-96
1213	cadherin	Cadherin domain	1.3e-06	24.8	2	143-235
1213	cadherin	Cadherin domain	1.3e-22	80.6	3	249-343
1213	cadherin	Cadherin domain	2.9e-14	51.4	4	361-448
1213	cadherin	Cadherin domain	1.7e-22	80.2	5	462-558
1213	cadherin	Cadherin domain	2.4e-10	37.8	6	597-667
1214	calreticulin	Calreticulin family	3.6e-221	715.1	1	21-315
1218	Alpha_L_fucos	Alpha-L-fucosidase	0.018	8.4	1	10-34
1221	Osteopontin	Osteopontin	1.4e-20	64.7	1	1-30
1221	Osteopontin	Osteopontin	2e-166	531.5	2	31-275
1222	serpin	Serpin (serine protease inhibitor)	2.4e-156	529.6	1	78-443
1223	ig	Immunoglobulin domain	0.00035	18.5	1	31-101
1223	ig	2/3 143 210 .. 7 52	1.3e-08	35.1	3	252-303
1225	HATPase_c	Histidine kinase-, DNA gyrase B-, and	3.8e-15	54.5	1	16-164
1225	DNA_gyraseB	DNA gyrase B	4.1e-57	199.9	1	210-370
1225	DNA_topoisolV	DNA gyrase/topoisomerase IV, subunit	1.3e-189	610.1	1	653-1120
1225	DUF188	Uncharacterized BCR, Yail/YqxD family	0.025	8.2	1	1095-1121

545
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1226	AMP-binding	AMP-binding enzyme	7.3e-84	288.8	1	105-539
1227	PCI	PCI domain	8.7e-10	36.4	1	46-101
1228	C1q	C1q domain	6e-45	159.5	1	73-202
1229	BTB	BTB/POZ domain	9.8e-14	51.0	1	9-69
1229	BTB	BTB/POZ domain	0.68	4.5	2	70-101
1230	ank	Ankyrin repeat	0.93	5.8	1	7-39
1230	ank	Ankyrin repeat	0.00039	17.9	2	40-85
1230	ank	Ankyrin repeat	2.1e-05	22.4	3	86-147
1230	ank	Ankyrin repeat	0.057	10.1	4	148-180
1230	ank	Ankyrin repeat	3.6e-10	39.5	5	181-213
1230	ank	Ankyrin repeat	7.2e-08	31.3	6	214-246
1230	ank	Ankyrin repeat	3.7e-06	25.1	7	247-279
1230	ank	Ankyrin repeat	1.7e-08	33.6	8	280-312
1230	ank	Ankyrin repeat	4.9e-07	28.3	9	313-346
1230	ank	Ankyrin repeat	0.00014	19.6	10	347-379
1230	ank	Ankyrin repeat	1.8e-07	29.9	11	380-412
1230	ank	Ankyrin repeat	0.038	10.8	12	413-437
1230	ank	Ankyrin repeat	2.5e-08	32.9	13	464-496
1230	ank	Ankyrin repeat	7.6e-08	31.2	14	497-529
1230	ank	Ankyrin repeat	2.2e-07	29.5	15	530-590
1230	ank	Ankyrin repeat	0.0048	14.0	16	591-613
1230	ank	Ankyrin repeat	0.0097	12.9	17	629-658
1230	ank	Ankyrin repeat	3.3e-06	25.3	18	659-691
1230	ank	Ankyrin repeat	2.3e-05	22.3	19	693-727
1230	ank	Ankyrin repeat	3.1e-09	36.2	20	729-761
1230	ank	Ankyrin repeat	0.00054	17.4	21	762-794
1230	ank	Ankyrin repeat	5.7e-06	24.5	22	795-827
1230	ank	Ankyrin repeat	9.6e-05	20.1	23	832-855
1230	ank	Ankyrin repeat	0.0013	16.0	24	864-892
1230	ank	Ankyrin repeat	1.7e-08	33.5	25	899-931
1230	ank	Ankyrin repeat	3.4e-06	25.3	26	932-965
1230	ank	Ankyrin repeat	0.001	16.4	27	966-998
1230	ank	Ankyrin repeat	2.5e-07	29.3	28	1006-1034
1231	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal do	1.5e-61	210.5	1	38-217
1231	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	7.3e-27	96.9	1	242-472
1232	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal do	1.5e-61	210.5	1	38-217
1232	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	1.3e-25	92.4	1	242-472
1233	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal do	1.5e-61	210.5	1	38-217
1233	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	2.6e-33	120.1	1	242-478
1234	DUF408	Domain of Unknown Function (DUF408)	7.8e-114	388.3	1	41-222
1237	ig	Immunoglobulin domain	5.3e-05	21.6	1	28-86
1237	ig	Immunoglobulin domain	2e-08	34.4	2	127-184
1237	ig	Immunoglobulin domain	6.2e-13	51.3	3	219-277
1237	fn3	Fibronectin type III domain	6.3e-20	71.0	1	299-385
1237	fn3	Fibronectin type III domain	8e-10	35.9	2	396-481
1238	Nuf2	Nuf2 family	3.3e-	356.3	1	1-148

546

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
			104			
1238	HR1	Hr1 repeat	0.099	7.1	1	187-214
1240	Sema	Sema domain	7.6e-178	601.0	1	59-477
1240	squash	Squash family serine protease inhibitor	0.033	5.8	1	512-534
1240	PSI	Plexin repeat	0.00097	13.3	1	514-543
1240	UreE_C	UreE urease accessory protein, C-terminal do	0.09	7.9	1	809-830
1243	rrm	RNA recognition motif. (a.k.a. RRM, RBD, or	0.0012	15.1	1	29-93
1247	Peptidase_M50	Peptidase family M50	0.0024	12.4	1	9-875
1247	C_tripleX	Cysteine rich repeat	0.00046	13.8	1	103-120
1247	EGF	EGF-like domain	0.089	9.2	1	105-135
1247	EGF	EGF-like domain	0.0047	13.8	2	148-178
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.0006	15.5	1	152-195
1247	EB	EB module	0.21	5.7	2	188-217
1247	EGF	EGF-like domain	3.2e-06	25.2	3	191-221
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.17	7.0	2	199-238
1247	EGF	EGF-like domain	8.6e-06	23.6	4	234-264
1247	DSL	Delta serrate ligand	0.024	8.9	4	250-264
1247	EGF	EGF-like domain	5.5e-06	24.3	5	277-307
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	1.2e-05	21.3	4	281-317
1247	EGF	EGF-like domain	0.00028	18.2	6	320-350
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.018	10.3	5	324-351
1247	DSL	Delta serrate ligand	0.15	6.3	6	353-364
1247	EGF	EGF-like domain	0.053	10.0	7	364-396
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.00011	18.0	6	368-406
1247	DSL	Delta serrate ligand	0.034	8.4	7	383-396
1247	DSL	Delta serrate ligand	0.76	4.0	8	397-409
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.68	4.9	7	413-456
1247	EGF	EGF-like domain	4.8e-05	20.9	9	415-439
1247	DSL	Delta serrate ligand	0.46	4.7	9	425-439
1247	EGF	EGF-like domain	1.2e-05	23.1	10	452-482
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.19	6.9	8	460-499
1247	DSL	Delta serrate ligand	0.21	5.8	10	469-482
1247	EB	EB module	0.89	3.8	4	492-525
1247	EGF	EGF-like domain	2.3e-05	22.1	11	495-525
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.011	11.1	9	502-542
1247	DSL	Delta serrate ligand	0.023	9.0	11	512-525
1247	EGF	EGF-like domain	0.03	10.8	12	538-568
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.0055	12.1	10	546-587
1247	DSL	Delta serrate ligand	0.0012	13.1	12	553-568
1247	EGF	EGF-like domain	0.0001	19.8	13	581-611
1247	EB	EB module	0.025	8.5	5	587-611
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.065	8.4	11	589-631
1247	DSL	Delta serrate ligand	0.5	4.6	14	612-624
1247	EGF	EGF-like domain	0.48	6.5	14	614-624
1247	EGF	EGF-like domain	0.041	10.4	15	631-669
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.00028	16.6	12	634-658
1247	EGF	EGF-like domain	0.00023	18.5	16	675-699
1247	DSL	15/20 647 656 .. 58 67	0.15	6.3	16	689-699
1247	EGF	EGF-like domain	7.7e-05	20.2	17	712-742
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.56	5.2	13	716-752
1247	DSL	15/20 647 656 .. 58 67	0.083	7.1	17	729-742

547

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1247	EGF	EGF-like domain	0.57	6.2	18	745-755
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	3.4e-05	19.8	14	759-802
1247	EB	EB module	0.079	7.0	6	760-785
1247	EGF	EGF-like domain	0.0048	13.7	19	761-785
1247	DSL	15/20 647 656 .. 58 67	0.16	6.2	19	772-785
1247	EGF	EGF-like domain	0.0057	13.5	20	798-828
1247	laminin_EGF	Laminin EGF-like (Domains III and V)	0.0035	12.8	15	805-830
1247	DSL	15/20 647 656 .. 58 67	0.44	4.8	20	809-828
1249	IBR	IBR domain	1.1e-05	19.4	1	74-104
1249	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.029	6.4	1	114-134
1249	IBR	IBR domain	0.029	8.4	2	132-164
1250	NC	NC domain	2.2e-47	167.3	1	172-253
1251	Aa_trans	Transmembrane amino acid transporter protein	1.9e-77	267.5	1	52-365
1252	Aa_trans	Transmembrane amino acid transporter protein	2.5e-76	263.7	1	45-354
1252	Aa_trans	Transmembrane amino acid transporter protein	2.7e-06	22.9	2	355-419
1254	FGF	Fibroblast growth factor	2.2e-40	137.9	1	42-166
1255	LRR	Leucine Rich Repeat	0.033	10.2	1	49-70
1255	LRR	Leucine Rich Repeat	0.21	7.5	2	71-92
1255	LRR	Leucine Rich Repeat	0.57	6.0	3	94-115
1255	LRR	Leucine Rich Repeat	0.46	6.3	4	116-140
1256	RPE65	Retinal pigment epithelial membrane protein	8.9e-59	199.4	1	60-416
1256	RPE65	Retinal pigment epithelial membrane protein	4.3e-27	91.2	2	462-579
1257	RPE65	Retinal pigment epithelial membrane protein	8.9e-59	199.4	1	42-398
1257	RPE65	Retinal pigment epithelial membrane protein	4.3e-27	91.2	2	444-561
1258	ig	Immunoglobulin domain	0.001	16.8	1	39-97
1258	ig	Immunoglobulin domain	1.2e-11	46.5	2	128-189
1260	DUF948	Bacterial protein of unknown function (DU	0.058	8.0	1	249-269
1261	serpin	Serpin (serine protease inhibitor)	3.2e-08	27.5	1	31-82
1261	serpin	Serpin (serine protease inhibitor)	1.2e-60	206.3	2	212-423
1262	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	3.7e-16	56.5	1	1-47
1263	arf	ADP-ribosylation factor family	5.1e-13	43.2	1	10-132
1264	PAP2	PAP2 superfamily	4.5e-15	54.4	1	106-241
1265	SRCR	Scavenger receptor cysteine-rich doma	2e-20	73.6	1	37-128
1265	SRCR	Scavenger receptor cysteine-rich doma	6e-28	100.2	2	136-227
1265	SRCR	Scavenger receptor cysteine-rich doma	6.6e-33	117.8	3	232-329
1265	Arthro_defensin	Arthropod defensin	0.0097	7.3	1	340-364
1265	SRCR	Scavenger receptor cysteine-rich doma	3.1e-15	55.3	4	360-459
1265	SRCR	Scavenger receptor cysteine-rich doma	7.6e-33	117.6	5	477-574
1266	SRCR	Scavenger receptor cysteine-rich doma	2e-20	73.6	1	37-128
1266	SRCR	Scavenger receptor cysteine-rich doma	6e-28	100.2	2	136-227
1266	SRCR	Scavenger receptor cysteine-rich doma	6.6e-33	117.8	3	232-329
1266	Arthro_defensin	Arthropod defensin	0.0097	7.3	1	340-364
1266	SRCR	Scavenger receptor cysteine-rich doma	3.1e-15	55.3	4	360-459
1266	SRCR	Scavenger receptor cysteine-rich doma	7.6e-33	117.6	5	477-574
1270	Armadillo_seg	Armadillo/beta-catenin-like repeat	2.7e-05	21.8	1	53-93
1270	Armadillo_seg	2/5 546 586 .. 1 41	0.11	9.0	5	691-716

548

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1273	Pneumo_att_G	Pneumovirinae attachment membrane glycop	0.098	4.7	1	57-70
1273	pkinese	Protein kinase domain	3e-77	266.8	1	103-387
1275	Pep_M12B_prop ep	Reprolysin family propeptide	3.2e-37	116.5	1	97-215
1275	Reprolysin	Reprolysin (M12B) family zinc metallo	1.1e-88	304.8	1	227-426
1275	Peptidase_M46	Pregnancy-associated plasma protein-A	0.056	5.5	1	362-372
1275	disintegrin	Disintegrin	1.7e-39	134.2	1	443-518
1275	EGF	EGF-like domain	0.0023	14.8	1	670-697
1276	FeoA	FeoA family	0.088	8.4	1	132-239
1277	ank	Ankyrin repeat	2.3e-05	22.3	1	301-339
1277	ank	Ankyrin repeat	9.5e-11	41.6	2	340-373
1277	Dehydratase_LU	Dehydratase large subunit	0.015	7.6	1	369-403
1278	Peptidase_M1	Peptidase family M1	7.1e-137	383.8	1	98-506
1284	Aa_trans	Transmembrane amino acid transporter	2.4e-30	110.9	1	4-397
1285	ARPF	Aromatic-Rich Protein Family	4.3e-09	31.3	1	74-190
1288	LRR	Leucine Rich Repeat	0.41	6.5	1	66-89
1288	LRR	Leucine Rich Repeat	0.0017	14.6	2	90-113
1288	LRR	Leucine Rich Repeat	0.76	5.6	3	114-137
1288	LRR	Leucine Rich Repeat	0.0013	14.9	4	138-161
1288	LRR	Leucine Rich Repeat	0.0043	13.1	5	163-186
1288	LRR	Leucine Rich Repeat	0.0088	12.1	6	187-210
1288	LRR	Leucine Rich Repeat	0.063	9.2	7	211-231
1288	LRRCT	Leucine rich repeat C-terminal domain	2.6e-10	32.7	1	252-297
1288	ig	Immunoglobulin domain	5.8e-09	36.4	1	314-372
1289	Huntingtin	Huntingtin	0.077	5.4	1	768-790
1290	LRRNT	Leucine rich repeat N-terminal domain	0.0011	14.5	1	32-59
1290	LRR	Leucine Rich Repeat	0.0059	12.7	1	61-84
1290	LRR	Leucine Rich Repeat	0.00021	17.6	2	85-108
1290	LRR	Leucine Rich Repeat	0.012	11.6	3	110-132
1290	LRRCT	Leucine rich repeat C-terminal domain	0.00014	15.2	1	131-144
1291	PH	PH domain	0.053	7.8	1	7-98
1291	DAGKc	Diacylglycerol kinase catalytic domain (pres	0.00081	14.7	1	90-177
1292	ig	Immunoglobulin domain	0.069	9.9	1	48-120
1292	ig	Immunoglobulin domain	8.1e-09	35.9	2	161-219
1293	ig	Immunoglobulin domain	0.069	9.9	1	48-120
1293	ig	Immunoglobulin domain	8.1e-09	35.9	2	161-219
1295	C1q	C1q domain	5.3e-49	173.0	1	72-198
1296	7tm_1	7 transmembrane receptor (rhodopsin famil	1.3e-08	24.4	1	49-108
1296	7tm_1	7 transmembrane receptor (rhodopsin famil	4.6e-31	91.7	2	109-332
1297	MED7	MED7 protein	0.0099	9.5	1	202-242
1297	CH	Calponin homology (CH) domain	2.7e-31	114.2	1	215-316
1297	CH	Calponin homology (CH) domain	3.7e-26	97.1	2	331-433
1297	UVR	UvrB/uvrC motif	0.0066	12.8	1	652-664
1297	spectrin	Spectrin repeat	0.007	11.5	1	793-852
1297	ACCA	Acetyl co-enzyme A carboxylase carboxy	0.017	10.3	1	832-873
1297	spectrin	Spectrin repeat	4.9e-05	18.9	2	922-973
1297	PolC_DP2	DNA polymerase II large subunit DP2	0.013	2.0	1	928-939
1297	DUF622	Protein of unknown function, DUF622	0.043	9.8	1	1313-

549
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						1341
1297	Myc-LZ	Myc leucine zipper domain	0.13	7.7	2	1313-1338
1297	spectrin	Spectrin repeat	0.38	5.5	3	1486-1512
1297	bZIP	1/3 644 674 .. 35 65	0.058	8.8	3	1698-1722
1297	Prefoldin	Prefoldin subunit	0.56	5.2	3	1709-1736
1297	M	M protein repeat	0.44	8.1	2	1939-1959
1297	ldh_C	lactate/malate dehydrogenase, alpha/be	0.35	5.2	2	2093-2118
1297	FTCD_C	Formiminotransferase-cyclodeaminase	0.029	9.2	1	2108-2146
1297	Laminin_II	Laminin Domain II	0.032	9.5	1	2152-2219
1297	Tropomyosin	Tropomyosin	0.019	8.9	1	2210-2251
1297	Pox_A_type_inc	2/7 1057 1069 .. 1 13	0.47	6.6	6	2364-2379
1297	Tropomyosin	Tropomyosin	0.72	3.2	2	2396-2425
1297	Pox_A_type_inc	2/7 1057 1069 .. 1 13	0.57	6.3	7	2399-2421
1297	Plectin	Plectin repeat	1e-19	74.9	2	2734-2778
1297	Plectin	Plectin repeat	8.3e-16	60.6	3	2808-2852
1297	CBM_14	Chitin binding Peritrophin-A domain	0.0038	11.3	1	2867-2884
1297	Plectin	Plectin repeat	2e-05	22.8	4	2907-2939
1297	Plectin	Plectin repeat	0.018	12.0	6	3012-3042
1297	Plectin	Plectin repeat	2.1e-20	77.4	7	3043-3087
1297	ECH	Enoyl-CoA hydratase/isomerase family	0.00096	14.0	1	3059-3080
1297	Plectin	Plectin repeat	0.083	9.6	8	3088-3118
1297	Plectin	Plectin repeat	1.3e-16	63.5	9	3119-3163
1297	Plectin	Plectin repeat	0.44	6.9	10	3169-3201
1298	MED7	MED7 protein	0.0099	9.5	1	202-242
1298	CH	Calponin homology (CH) domain	2.3e-29	107.8	1	215-328
1298	CH	Calponin homology (CH) domain	3.7e-26	97.1	2	343-445
1298	UVR	UvrB/uvrC motif	0.0066	12.8	1	664-676
1298	spectrin	Spectrin repeat	0.007	11.5	1	805-864
1298	ACCA	Acetyl co-enzyme A carboxylase carboxy	0.017	10.3	1	844-885
1298	spectrin	Spectrin repeat	4.9e-05	18.9	2	934-985
1298	PolC_DP2	DNA polymerase II large subunit DP2	0.013	2.0	1	940-951
1298	DUF622	Protein of unknown function, DUF622	0.043	9.8	1	1325-

550
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
						1353
1298	Myc-LZ	Myc leucine zipper domain	0.13	7.7	2	1325-1350
1298	spectrin	Spectrin repeat	0.38	5.5	3	1498-1524
1298	bZIP	1/3 656 686 .. 35 65	0.058	8.8	3	1710-1734
1298	Prefoldin	Prefoldin subunit	0.56	5.2	3	1721-1748
1298	M	M protein repeat	0.44	8.1	2	1951-1971
1298	ldh_C	lactate/malate dehydrogenase, alpha/be	0.35	5.2	2	2105-2130
1298	FTCD_C	Formiminotransferase-cyclodeaminase	0.029	9.2	1	2120-2158
1298	Laminin_II	Laminin Domain II	0.032	9.5	1	2164-2231
1298	Tropomyosin	Tropomyosin	0.019	8.9	1	2222-2263
1298	Pox_A_type_inc	2/7 1069 1081 .. 1 13	0.47	6.6	6	2376-2391
1298	Tropomyosin	Tropomyosin	0.72	3.2	2	2408-2437
1298	Pox_A_type_inc	2/7 1069 1081 .. 1 13	0.57	6.3	7	2411-2433
1298	Plectin	Plectin repeat	1e-19	74.9	2	2746-2790
1298	Plectin	Plectin repeat	8.3e-16	60.6	3	2820-2864
1298	CBM_14	Chitin binding Peritrophin-A domain	0.0038	11.3	1	2879-2896
1298	Plectin	Plectin repeat	2e-05	22.8	4	2919-2951
1298	Plectin	Plectin repeat	0.018	12.0	6	3024-3054
1298	Plectin	Plectin repeat	2.1e-20	77.4	7	3055-3099
1298	ECH	Enoyl-CoA hydratase/isomerase family	0.00096	14.0	1	3071-3092
1298	Plectin	Plectin repeat	0.083	9.6	8	3100-3130
1298	Plectin	Plectin repeat	1.3e-16	63.5	9	3131-3175
1298	Plectin	Plectin repeat	0.44	6.9	10	3181-3213
1304	DUF544	Protein of unknown function (DUF544)	5.8e-80	275.8	1	157-282
1305	DUF544	Protein of unknown function (DUF544)	5.8e-80	275.8	1	272-397
1306	ig	Immunoglobulin domain	2.2e-08	34.3	1	26-93
1306	ig	Immunoglobulin domain	2.5e-06	26.5	2	132-191
1306	MAM	MAM domain	6.9e-72	249.0	1	422-595
1308	APH	Phosphotransferase enzyme family	2.9e-42	150.6	1	40-256
1308	Acyl-CoA_dh_M	Acyl-CoA dehydrogenase, middle domain	0.00024	17.0	1	505-585
1308	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal doma	6.7e-50	175.9	1	618-769

551
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1309	APH	Phosphotransferase enzyme family	7.7e-32	116.0	1	80-238
1309	Acyl-CoA_dh_M	Acyl-CoA dehydrogenase, middle domain	0.00024	17.0	1	487-567
1309	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal doma	6.7e-50	175.9	1	600-751
1310	Cation_efflux	Cation efflux family	3e-09	34.4	1	69-145
1311	CaMBD	Calmodulin binding domain	0.074	7.8	1	716-732
1311	IQ	IQ calmodulin-binding motif	1.3e-05	22.1	2	738-758
1312	SAM	SAM domain (Sterile alpha motif)	0.00073	15.4	1	304-369
1312	SAM	SAM domain (Sterile alpha motif)	2.2e-10	38.2	2	382-446
1312	SAM	SAM domain (Sterile alpha motif)	0.06	8.7	3	470-499
1313	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	7.1e-25	70.3	1	80-126
1313	Herpes_UL49_5	Herpesvirus UL49.5 envelope/tegument pr	0.082	7.4	1	147-170
1314	DUF692	Protein of unknown function (DUF692)	0.088	6.2	1	1703-1722
1314	HECT	HECT-domain (ubiquitin-transferase)	1.9e-196	662.8	1	2002-2309
1314	V-ATPase_C	V-ATPase subunit C	0.032	7.2	1	2185-2213
1315	PAP2	PAP2 superfamily	1.5e-25	91.6	1	66-218
1316	PAP2	PAP2 superfamily	5.1e-30	107.5	1	98-236
1317	ig	Immunoglobulin domain	8.1e-09	35.9	1	41-116
1321	LRRNT	Leucine rich repeat N-terminal domain	1.3e-06	24.2	1	115-143
1321	LRR	Leucine Rich Repeat	0.098	8.6	1	145-168
1321	LRR	Leucine Rich Repeat	8.2e-06	22.3	2	169-194
1321	FNIP	FNIP Repeat	0.36	6.7	1	195-225
1321	LRR	Leucine Rich Repeat	0.63	5.9	3	195-207
1321	LRR	Leucine Rich Repeat	0.0026	13.9	4	240-265
1321	LRR	Leucine Rich Repeat	0.018	11.1	5	266-285
1321	LRR	Leucine Rich Repeat	0.00014	18.1	6	287-310
1321	LRR	Leucine Rich Repeat	0.00013	18.3	7	311-336
1321	LRR	Leucine Rich Repeat	0.00015	18.1	8	337-356
1321	LRR	Leucine Rich Repeat	0.22	7.4	9	358-381
1321	LRR	Leucine Rich Repeat	0.002	14.3	10	382-407
1321	LRR	Leucine Rich Repeat	0.022	10.7	11	408-427
1321	LRR	Leucine Rich Repeat	0.00025	17.3	12	453-478
1321	LRR	Leucine Rich Repeat	0.00049	16.4	13	479-498
1321	LRR	Leucine Rich Repeat	0.13	8.1	15	524-549
1321	LRR	Leucine Rich Repeat	0.00025	17.3	16	550-569
1321	LRR	Leucine Rich Repeat	5.2e-05	19.6	17	571-594
1321	LRR	Leucine Rich Repeat	0.37	6.6	18	595-620
1322	ig	Immunoglobulin domain	0.26	7.8	1	50-117
1322	ig	Immunoglobulin domain	0.00049	18.0	2	157-215
1322	ig	Immunoglobulin domain	2.8e-09	37.6	3	267-321
1323	ig	Immunoglobulin domain	0.24	7.9	1	50-117
1323	ig	Immunoglobulin domain	0.00049	18.0	2	157-215
1323	ig	Immunoglobulin domain	0.00077	17.2	3	267-303
1324	tsp_1	Thrombospondin type 1 domain	2.9e-07	25.9	1	37-81
1325	Guanylin	Guanylin precursor	0.00035	9.9	1	1-24
1325	Apo-CII	Apolipoprotein C-II	9.1e-43	152.3	1	23-99
1326	Guanylin	Guanylin precursor	0.00035	9.9	1	1-24
1326	Apo-CII	Apolipoprotein C-II	9.1e-43	152.3	1	23-99
1328	SRCR	Scavenger receptor cysteine-rich	6.5e-37	131.9	1	14-111

552
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
		domain				
1328	SRCR	Scavenger receptor cysteine-rich domain	1.2e-34	123.9	2	188-285
1328	SRCR	Scavenger receptor cysteine-rich domain	4.7e-37	132.4	3	300-397
1328	SRCR	Scavenger receptor cysteine-rich domain	1.5e-35	127.1	4	405-503
1328	DUF159	Uncharacterised ACR, COG2135	0.092	3.7	1	565-587
1328	SRCR	Scavenger receptor cysteine-rich domain	1.8e-27	98.6	5	638-729
1329	ig	Immunoglobulin domain	0.81	5.9	1	37-84
1329	ig	Immunoglobulin domain	0.051	10.4	2	113-165
1331	efhand	EF hand	0.025	12.1	1	12-40
1331	efhand	EF hand	0.97	6.2	2	59-76
1331	efhand	EF hand	0.041	11.2	3	85-113
1333	wnt	wnt family	6.9e-240	694.6	1	40-365
1335	7tm_1	7 transmembrane receptor (rhodopsin family)	8.8e-18	51.9	1	8-75
1336	SAP	SAP domain	3.8e-07	29.1	1	11-45
1336	zf-MIZ	MIZ zinc finger	4.1e-41	120.1	1	323-375
1337	FA_desaturase	Fatty acid desaturase	1.2e-76	264.7	1	71-296
1338	cystatin	Cystatin domain	0.074	6.5	1	25-45
1340	actin	Actin	5.4e-67	221.4	1	4-362
1340	E1_N	E1 Protein, N terminal domain	0.08	6.5	1	149-158
1341	ion_trans	Ion transport protein	0.007	10.9	1	114-168
1341	ion_trans	Ion transport protein	5e-05	18.6	2	211-302
1343	ig	Immunoglobulin domain	6.1e-06	25.1	1	124-182
1343	ig	Immunoglobulin domain	2.2e-06	26.8	2	224-281
1343	ig	Immunoglobulin domain	7.6e-08	32.2	3	316-372
1343	fn3	Fibronectin type III domain	2.8e-16	58.3	1	394-480
1343	fn3	Fibronectin type III domain	6.6e-17	60.5	2	492-578
1343	fn3	Fibronectin type III domain	0.013	10.8	3	598-654
1344	DUF84	Protein of unknown function DUF84	0.098	5.9	1	8-22
1344	ig	Immunoglobulin domain	3e-07	30.0	1	53-110
1344	ig	Immunoglobulin domain	1.8e-07	30.9	2	150-216
1344	ig	Immunoglobulin domain	2.9e-08	33.8	3	255-310
1344	ig	Immunoglobulin domain	4.6e-07	29.3	4	350-417
1344	ig	Immunoglobulin domain	1.1e-07	31.6	5	456-516
1344	ig	Immunoglobulin domain	8.8e-05	20.8	6	553-617
1344	MAM	MAM domain	6.7e-77	265.6	1	753-918
1345	kazal	Kazal-type serine protease inhibitor domain	7.7e-06	25.8	1	121-168
1345	ig	Immunoglobulin domain	1.2e-06	27.7	1	186-255
1346	RNA_helicase	RNA helicase	0.031	7.9	1	82-109
1346	ATP-bind	Conserved hypothetical ATP binding pr	0.055	7.3	1	87-100
1348	ig	Immunoglobulin domain	8.5e-07	28.3	1	61-120
1348	ig	Immunoglobulin domain	0.00026	19.0	2	155-214
1348	ig	Immunoglobulin domain	4.7e-08	33.0	3	258-315
1348	ig	Immunoglobulin domain	2.3e-05	23.0	4	348-404
1348	ig	Immunoglobulin domain	4.6e-09	36.8	5	440-497
1348	ig	Immunoglobulin domain	8.8e-07	28.3	6	530-596
1348	fn3	Fibronectin type III domain	5.2e-20	71.3	1	615-704
1348	fn3	Fibronectin type III domain	0.0015	14.1	2	717-807

553

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1348	fn3	Fibronectin type III domain	8.9e-14	49.6	3	819-907
1348	fn3	Fibronectin type III domain	0.00019	17.2	4	919-1002
1350	serpin	Serpin (serine protease inhibitor)	5.1e-197	664.7	1	45-378
1350	serpin	Serpin (serine protease inhibitor)	8e-09	29.6	2	379-402
1352	DREV	DREV methyltransferase	7.3e-233	680.7	1	56-317
1353	CARD	Caspase recruitment domain	2.6e-33	119.8	1	2-91
1355	ank	Ankyrin repeat	1.8e-07	29.9	2	64-96
1355	ank	Ankyrin repeat	1.6e-06	26.4	3	97-129
1355	ank	Ankyrin repeat	3.8e-07	28.7	4	130-162
1355	ank	Ankyrin repeat	0.00011	19.9	5	163-195
1355	ank	Ankyrin repeat	0.00012	19.8	6	196-228
1356	pkinase	Protein kinase domain	3.5e-64	223.4	1	221-479
1356	Aldolase	KDPG and KHG aldolase	0.038	7.4	1	868-891
1357	pkinase	Protein kinase domain	2.8e-05	18.9	1	43-72
1357	Aldolase	KDPG and KHG aldolase	0.038	7.4	1	461-484
1358	7tm_1	7 transmembrane receptor (rhodopsin family)	2.6e-13	38.4	1	1-59
1359	tRNA-synt_1	tRNA synthetases class I (I, L, M and V)	0.00037	12.8	1	53-115
1359	tRNA-synt_1e	tRNA synthetases class I (C)	0.0002	14.0	1	345-375
1359	tRNA-synt_1	tRNA synthetases class I (I, L, M and V)	2.4e-07	23.7	2	345-383
1360	MHC_II_beta	Class II histocompatibility antigen, beta	1.4e-43	149.3	1	42-117
1363	ig	Immunoglobulin domain	0.86	5.8	1	12-69
1363	ig	Immunoglobulin domain	0.17	8.4	2	139-200
1363	ig	Immunoglobulin domain	0.00066	17.5	3	236-294
1363	ig	Immunoglobulin domain	7.9e-06	24.7	4	344-398
1364	fn3	Fibronectin type III domain	0.0032	12.9	1	35-125
1365	IL1	Interleukin-1 / 18	5.4e-31	110.6	1	11-155
1366	A2M_N	Alpha-2-macroglobulin family N-terminal regi	1.5e-92	317.7	1	6-613
1366	A2M	Alpha-2-macroglobulin family	3.6e-211	711.7	1	722-1449
1367	ABC_membrane	ABC transporter transmembrane region	1.7e-07	28.5	1	1-70
1368	UPAR_LY6	u-PAR/Ly-6 domain	2.6e-37	134.1	1	27-106
1967	DUF99	Protein of unknown function DUF99	0.06	5.8	1	3-26
1967	hormone	Somatotropin hormone family	1.6e-55	156.0	1	29-141
1968	DUF99	Protein of unknown function DUF99	0.06	5.8	1	3-26
1968	hormone	Somatotropin hormone family	1.6e-55	156.0	1	29-141
1969	DUF99	Protein of unknown function DUF99	0.06	5.8	1	3-26
1969	hormone	Somatotropin hormone family	1.6e-55	156.0	1	29-141
1970	DUF99	Protein of unknown function DUF99	0.06	5.8	1	3-26
1970	hormone	Somatotropin hormone family	1.6e-55	156.0	1	29-141
1971	serpin	Serpin (serine protease inhibitor)	5.1e-83	282.6	1	83-449
1972	PI-PLC-X	Phosphatidylinositol-specific phospholipase	3.8e-14	50.6	1	1-33
1973	Lipase_3	Lipase (class 3)	1.7e-17	62.0	1	399-538
1976	DUF846	Eukaryotic protein of unknown function (DUF8	0.0091	7.9	1	79-109
1977	Monooxygenase	Monooxygenase	3.8e-12	44.1	1	215-313
1977	Monooxygenase	Monooxygenase	1.7e-15	56.1	2	358-443

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
1980	zf-AN1	AN1-like Zinc finger	0.032	10.1	1	59-98
1980	zf-AN1	AN1-like Zinc finger	9.2e-06	22.6	2	149-181
1981	CRAL_TRIO	CRAL/TRIO domain	0.037	7.8	1	10-38
1982	Rhomboid	Rhomboid family	3.9e-32	116.9	1	128-282
1984	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal do	4.5e-38	130.4	1	33-191
1984	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	8.3e-14	49.9	1	253-456
1987	DUF572	Family of unknown function (DUF572)	3.5e-37	133.7	1	1-61
1987	DUF572	Family of unknown function (DUF572)	5e-23	84.4	2	91-149
1988	Collagen	Collagen triple helix repeat (20 copi	3.7e-11	44.2	1	1-51
1988	Collagen	Collagen triple helix repeat (20 copi	6.6e-11	43.2	2	60-115
1988	Collagen	Collagen triple helix repeat (20 copi	3.9e-13	51.6	3	116-175
1988	Collagen	Collagen triple helix repeat (20 copi	0.0069	13.1	4	178-195
1988	Collagen	Collagen triple helix repeat (20 copi	0.0001	20.0	5	199-230
1988	Collagen	Collagen triple helix repeat (20 copi	4e-09	36.5	6	239-298
1988	Collagen	Collagen triple helix repeat (20 copi	1.9e-13	52.8	7	302-355
1988	Collagen	Collagen triple helix repeat (20 copi	7.1e-06	24.3	8	362-395
1988	Collagen	Collagen triple helix repeat (20 copi	0.0012	16.0	9	396-444
1988	C4	C-terminal tandem repeated domain in	2e-69	240.8	1	450-557
1988	C4	C-terminal tandem repeated domain in	1.3e-77	268.0	2	558-672
1989	ldl_recept_b	Low-density lipoprotein receptor repeat	7.3e-10	34.9	1	56-97
1989	ldl_recept_b	Low-density lipoprotein receptor repeat	2.7e-07	26.4	2	99-141
1989	ldl_recept_b	Low-density lipoprotein receptor repeat	3.2e-07	26.2	3	143-185
1990	ldl_recept_b	Low-density lipoprotein receptor repeat	7.3e-10	34.9	1	56-97
1990	ldl_recept_b	Low-density lipoprotein receptor repeat	2.7e-07	26.4	2	99-141
1990	ldl_recept_b	Low-density lipoprotein receptor repeat	3.2e-07	26.2	3	143-185
1991	DUF846	Eukaryotic protein of unknown function (DUF8	0.00016	13.3	1	76-106
1992	cadherin	Cadherin domain	2.1e-10	38.0	1	9-105
1992	cadherin	Cadherin domain	1.4e-28	101.4	2	119-210
1993	cadherin	Cadherin domain	2.1e-10	38.0	1	9-105
1993	cadherin	Cadherin domain	1.4e-28	101.4	2	119-210
1995	V1R	Vomeronasal organ pheromone receptor family,	3.8e-08	27.0	1	4-36
1998	ig	Immunoglobulin domain	2.1e-09	38.1	1	18-76
1998	ig	Immunoglobulin domain	7.9e-09	35.9	2	121-179
1998	ig	Immunoglobulin domain	0.00014	20.0	3	216-274
1998	ig	Immunoglobulin domain	7.1e-09	36.1	4	308-366
1998	ig	Immunoglobulin domain	1.7e-10	42.2	5	403-461
1999	SPRY	SPRY domain	1.8e-30	107.5	1	148-277
1999	SRP54	SRP54-type protein, GTPase domain	0.0091	11.6	1	310-325
1999	AAA	ATPase family associated with various cellul	0.098	5.8	1	313-325
2000	ABC_tran	ABC transporter	2.5e-43	146.2	1	118-301
2002	Acyl-CoA_dh_M	Acyl-CoA dehydrogenase, middle domain	0.0071	11.7	1	99-136
2002	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal doma	6.7e-50	175.9	1	415-566
2003	C_tripleX	Cysteine rich repeat	2e-05	17.8	1	76-93
2003	EGF	EGF-like domain	8.7e-06	23.6	2	115-143
2003	TIL	Trypsin Inhibitor like cysteine rich domain	0.0035	11.0	1	134-155
2003	EGF	EGF-like domain	7.5e-05	20.2	3	155-189

555

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2003	TIL	Trypsin Inhibitor like cysteine rich domain	0.26	5.1	2	168-195
2003	EGF	EGF-like domain	4.4e-05	21.1	4	195-228
2003	EGF	EGF-like domain	9.7e-09	34.3	5	240-275
2003	MAM	MAM domain	9.2e-38	135.6	1	421-566
2004	NHL	NHL repeat	1.1e-10	42.4	1	8-35
2004	NHL	NHL repeat	2.5e-09	37.6	2	55-82
2004	NHL	NHL repeat	7.8e-11	43.0	3	102-129
2005	FCH	Fes/CIP4 homology domain	0.026	10.3	1	310-350
2005	DAG_PE-bind	Phorbol esters/diacylglycerol binding dom	2.8e-05	21.7	1	738-776
2005	RhoGAP	RhoGAP domain	3.9e-68	231.7	1	804-976
2006	CN_hydrolase	Carbon-nitrogen hydrolase	4.5e-07	26.2	2	117-206
2007	tsp_1	Thrombospondin type 1 domain	0.054	8.4	1	5-23
2008	Adaptin_N	Adaptin N terminal region	7.5e-09	29.6	1	1-51
2008	Alpha_adaptinC2	Adaptin C-terminal domain	4.4e-38	126.8	1	183-296
2008	Alpha_adaptin_C	Alpha adaptin AP2, C-terminal domain	1.6e-113	334.2	1	302-414
2009	ig	Immunoglobulin domain	0.0045	14.4	1	42-129
2009	ig	Immunoglobulin domain	0.19	8.3	2	179-272
2009	ig	Immunoglobulin domain	9.7e-05	20.6	3	319-408
2009	ig	Immunoglobulin domain	0.00014	20.0	4	455-546
2010	ig	Immunoglobulin domain	0.0045	14.4	1	42-129
2010	ig	Immunoglobulin domain	0.19	8.3	2	179-272
2010	ig	Immunoglobulin domain	9.7e-05	20.6	3	319-408
2010	ig	Immunoglobulin domain	0.00014	20.0	4	455-546
2011	ig	Immunoglobulin domain	0.0045	14.4	1	42-129
2011	ig	Immunoglobulin domain	0.19	8.3	2	179-272
2011	ig	Immunoglobulin domain	9.7e-05	20.6	3	319-408
2011	ig	Immunoglobulin domain	0.00014	20.0	4	455-546
2012	ig	Immunoglobulin domain	0.0045	14.4	1	42-129
2012	ig	Immunoglobulin domain	0.19	8.3	2	179-272
2012	ig	Immunoglobulin domain	9.7e-05	20.6	3	319-408
2012	ig	Immunoglobulin domain	0.00014	20.0	4	455-546
2016	TFA	Transcription elongation factor A, SII-r	3.4e-23	87.2	1	148-283
2018	cadherin	Cadherin domain	8e-13	46.4	1	1-49
2018	cadherin	Cadherin domain	4.5e-09	33.4	2	76-120
2019	PGM_PMM	Phosphoglucomutase/phosphomannomutase, C-ter	0.041	9.3	1	347-389
2020	MACPF	MAC/Perforin domain	0.00017	15.5	1	132-164
2021	KRAB	KRAB box	6.9e-24	88.6	1	54-94
2022	KRAB	KRAB box	6.9e-24	88.6	1	54-94
2023	EMP24_GP25L	emp24/gp25L/p24 family	1.9e-15	55.4	1	17-78
2024	acid_phosphat	Histidine acid phosphatase	7.9e-159	537.8	1	35-375
2026	KRAB	KRAB box	1.1e-20	77.0	1	132-172
2026	zf-C2H2	Zinc finger, C2H2 type	3.7e-07	33.4	1	485-507
2026	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.54	2.9	1	500-518
2026	zf-C2H2	Zinc finger, C2H2 type	1.3e-05	27.2	2	513-535
2026	zf-C2H2	Zinc finger, C2H2 type	3.4e-08	37.4	3	543-565
2026	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.032	6.3	2	558-576
2026	zf-C2H2	Zinc finger, C2H2 type	5.7e-06	28.6	4	571-593
2027	Vps16_N	Vps16, N-terminal region	2.3e-107	366.9	1	1-165

556
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2027	Vps16_C	Vps16, C-terminal region	5e-203	684.6	1	262-580
2028	LRRNT	Leucine rich repeat N-terminal domain	0.0011	14.5	1	11-40
2029	A2M	Alpha-2-macroglobulin family	6.3e-23	75.5	1	4-86
2031	fn3	Fibronectin type III domain	4.9e-08	29.7	1	3-33
2031	fn3	Fibronectin type III domain	9.1e-08	28.7	2	46-136
2032	LRR	Leucine Rich Repeat	0.021	10.8	1	2-25
2032	LRR	Leucine Rich Repeat	3e-05	20.4	2	26-49
2032	LRR	Leucine Rich Repeat	0.00019	17.8	3	50-73
2032	LRR	Leucine Rich Repeat	0.16	7.8	4	74-94
2032	LRRCT	Leucine rich repeat C-terminal domain	2.2e-05	17.6	1	118-132
2033	LRR	Leucine Rich Repeat	0.021	10.8	1	2-25
2033	LRR	Leucine Rich Repeat	3e-05	20.4	2	26-49
2033	LRR	Leucine Rich Repeat	0.00019	17.8	3	50-73
2033	LRR	Leucine Rich Repeat	0.16	7.8	4	74-94
2033	LRRCT	Leucine rich repeat C-terminal domain	2.2e-05	17.6	1	118-132
2034	EGF	EGF-like domain	0.76	5.8	1	135-157
2034	SEA	SEA domain	4.9e-06	22.1	1	192-261
2034	ig	Immunoglobulin domain	9.8e-07	28.1	1	310-376
2034	ig	Immunoglobulin domain	0.33	7.4	2	509-571
2034	GPS	Latrophilin/CL-1-like GPS domain	2e-14	54.5	1	975-1027
2034	7tm_2	7 transmembrane receptor (Secretin family)	2.8e-20	71.1	2	1086-1298
2035	TFIIS	Transcription factor S-II (TFIIS)	0.019	10.6	1	21-31
2035	zf-C2H2	Zinc finger, C2H2 type	3.1e-06	29.7	2	21-43
2035	zf-C2H2	Zinc finger, C2H2 type	2.7e-07	33.9	3	49-71
2035	zf-BED	BED zinc finger	0.63	4.8	1	50-72
2035	XPA_N	2/4 46 56 .. 1 11	0.22	7.0	3	74-86
2035	zf-C2H2	Zinc finger, C2H2 type	8.8e-08	35.9	4	77-99
2035	TFIIS	Transcription factor S-II (TFIIS)	0.036	9.7	4	105-115
2035	zf-C2H2	Zinc finger, C2H2 type	0.0096	15.6	5	105-120
2038	zf-C2H2	Zinc finger, C2H2 type	0.0099	15.5	1	197-220
2039	FHA	FHA domain	0.024	11.6	1	45-110
2039	HIT	HIT domain	0.013	8.5	1	201-226
2039	zf-C2H2	Zinc finger, C2H2 type	0.026	13.9	1	337-359
2040	FHA	FHA domain	0.024	11.6	1	45-110
2040	HIT	HIT domain	0.013	8.5	1	201-226
2040	zf-C2H2	Zinc finger, C2H2 type	0.026	13.9	1	337-359
2041	FHA	FHA domain	0.024	11.6	1	45-110
2041	HIT	HIT domain	0.013	8.5	1	201-226
2041	zf-C2H2	Zinc finger, C2H2 type	0.026	13.9	1	337-359
2042	Cwf_Cwc_15	Cwf15/Cwc15 cell cycle control protei	8.6e-161	544.3	1	2-230
2043	SRCR	Scavenger receptor cysteine-rich domain	6.5e-15	54.2	1	8-113
2043	Lysyl_oxidase	Lysyl oxidase	1.9e-140	476.7	1	117-286
2045	WD40	WD domain, G-beta repeat	0.5	6.4	2	192-217
2045	WD40	WD domain, G-beta repeat	5.2e-06	23.8	3	248-274
2045	DUF130	Domain of unknown function DUF130	0.074	5.9	1	264-278
2045	WD40	WD domain, G-beta repeat	0.35	7.0	4	397-424
2048	CTP_transf_1	Cytidyltransferase family	4.9e-124	422.2	1	86-417
2049	CBM_20	Starch binding domain	0.078	8.5	1	14-33

557

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2049	WD40	WD domain, G-beta repeat	3.9e-08	31.2	1	93-131
2052	7tm_3	7 transmembrane receptor (metabotropic gluta	5.8e-06	21.5	1	112-153
2052	KdgT	2-keto-3-deoxygluconate permease	0.068	7.2	1	200-226
2052	7tm_3	7 transmembrane receptor (metabotropic gluta	7.8e-05	17.4	2	221-305
2054	HesB-like	HesB-like domain	2.8e-41	132.5	1	52-154
2055	ig	Immunoglobulin domain	0.032	11.2	1	37-59
2055	ig	Immunoglobulin domain	0.00033	18.6	2	98-157
2056	Mpv17_PMP22	Mpv17 / PMP22 family	8e-14	51.5	1	101-163
2058	Collagen	Collagen triple helix repeat (20 copies)	0.013	12.1	2	17-38
2058	Collagen	Collagen triple helix repeat (20 copies)	2.5e-07	29.8	3	40-79
2058	vwa	von Willebrand factor type A domain	3.2e-13	42.1	1	108-156
2059	Sterol_desat	Sterol desaturase	8.6e-41	138.1	1	1-139
2060	ig	Immunoglobulin domain	0.27	7.7	1	8-26
2060	ig	Immunoglobulin domain	5.2e-08	32.9	2	97-158
2061	RNA_helicase	RNA helicase	0.00029	15.0	1	40-63
2061	AAA	ATPase family associated with various ce	0.00038	13.8	1	42-58
2061	NACHT	NACHT domain	0.0022	12.0	1	44-66
2061	ADK	Adenylate kinase	2.2e-05	19.0	1	77-124
2064	UDPGT	UDP-glucuronosyl and UDP-glucosyl transferas	9.7e-34	118.7	1	1-63
2065	TRAPP_Bet3	Transport protein particle (TRAPP) compone	9e-70	242.0	1	18-171
2066	DUF846	Eukaryotic protein of unknown function (DUF8	0.013	7.4	1	83-101
2068	ig	Immunoglobulin domain	0.0042	14.5	1	33-110
2068	FliL	Flagellar basal body-associated protein FliL	0.029	9.2	1	170-203
2068	DcuC	C4-dicarboxylate anaerobic carrier	0.044	7.9	1	174-193
2069	ig	Immunoglobulin domain	0.0042	14.5	1	33-110
2069	FliL	Flagellar basal body-associated protein FliL	0.029	9.2	1	170-203
2069	DcuC	C4-dicarboxylate anaerobic carrier	0.044	7.9	1	174-193
2070	ig	Immunoglobulin domain	0.0042	14.5	1	33-110
2070	FliL	Flagellar basal body-associated protein FliL	0.029	9.2	1	170-203
2070	DcuC	C4-dicarboxylate anaerobic carrier	0.044	7.9	1	174-193
2071	PH	PH domain	1.9e-21	72.0	1	75-173
2072	Ifi-6-16	Interferon-induced 6-16 family	3.7e-46	159.7	1	41-123
2073	Ifi-6-16	Interferon-induced 6-16 family	3.7e-46	159.7	1	41-123
2074	Ribosomal_L34e	Ribosomal protein L34e	3.5e-72	232.6	1	12-110
2075	CDC50	LEM3 (ligand-effect modulator 3) family / CD	0.049	6.6	1	90-117
2077	EGF	EGF-like domain	0.0019	15.2	1	60-95
2078	EGF	EGF-like domain	0.0019	15.2	1	60-95
2079	EGF	EGF-like domain	0.0019	15.2	1	60-95
2080	ig	Immunoglobulin domain	4.9e-06	25.4	1	109-171
2081	Monooxygenase	Monooxygenase	0.0069	10.9	1	593-611
2081	ras	Ras family	7.2e-10	33.6	1	924-967
2082	Alpha_adaptin_C	Alpha adaptin AP2, C-terminal domain	0.061	5.2	1	97-109
2082	MHC_I	Class I Histocompatibility antigen, d	0.00048	14.9	2	125-210
2083	ig	Immunoglobulin domain	4.1e-05	22.0	1	10-78

558
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2083	ig	Immunoglobulin domain	3.7e-10	40.9	2	113-172
2083	ig	Immunoglobulin domain	0.0018	15.9	3	211-272
2083	ig	Immunoglobulin domain	3.7e-08	33.4	4	309-370
2083	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	326-382
2083	OapA	Opacity-associated protein A	0.44	2.4	1	335-357
2083	ig	Immunoglobulin domain	0.0012	16.6	5	404-465
2083	ig	Immunoglobulin domain	7.7e-07	28.5	6	500-564
2084	ig	Immunoglobulin domain	4.1e-05	22.0	1	10-78
2084	ig	Immunoglobulin domain	3.7e-10	40.9	2	113-172
2084	ig	Immunoglobulin domain	0.0018	15.9	3	211-272
2084	ig	Immunoglobulin domain	3.7e-08	33.4	4	309-370
2084	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	326-382
2084	OapA	Opacity-associated protein A	0.44	2.4	1	335-357
2084	ig	Immunoglobulin domain	0.0012	16.6	5	404-465
2084	ig	Immunoglobulin domain	7.7e-07	28.5	6	500-564
2085	ig	Immunoglobulin domain	4.1e-05	22.0	1	10-78
2085	ig	Immunoglobulin domain	3.7e-10	40.9	2	113-172
2085	ig	Immunoglobulin domain	0.0018	15.9	3	211-272
2085	ig	Immunoglobulin domain	3.7e-08	33.4	4	309-370
2085	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	326-382
2085	OapA	Opacity-associated protein A	0.44	2.4	1	335-357
2085	ig	Immunoglobulin domain	0.0012	16.6	5	404-465
2085	ig	Immunoglobulin domain	7.7e-07	28.5	6	500-564
2086	P53	P53	3.5e-09	33.8	1	7-32
2087	Apolipoprotein	Apolipoprotein A1/A4/E family	2.3e-11	42.3	1	93-168
2087	DUF260	Protein of unknown function DUF260	0.64	3.5	1	94-107
2087	Adeno_PIX	Adenovirus hexon-associated protein (0.49	4.4	1	95-110
2087	BcrAD_BadFG	BadF/BadG/BcrA/BcrD ATPase family	0.12	6.2	1	134-180
2087	Apolipoprotein	Apolipoprotein A1/A4/E family	0.011	10.5	2	172-258
2087	MM_CoA_mutase	Methylmalonyl-CoA mutase	0.84	1.9	1	264-306
2088	Apolipoprotein	Apolipoprotein A1/A4/E family	2.3e-11	42.3	1	93-168
2088	DUF260	Protein of unknown function DUF260	0.64	3.5	1	94-107
2088	Adeno_PIX	Adenovirus hexon-associated protein (0.49	4.4	1	95-110
2088	BcrAD_BadFG	BadF/BadG/BcrA/BcrD ATPase family	0.12	6.2	1	134-180
2088	Apolipoprotein	Apolipoprotein A1/A4/E family	0.011	10.5	2	172-258
2088	MM_CoA_mutase	Methylmalonyl-CoA mutase	0.84	1.9	1	264-306
2089	DUF717	Protein of unknown function (DUF717)	1	4.0	1	68-80
2089	MHC_I	Class I Histocompatibility antigen, d	0.69	3.7	1	185-198
2090	Pox_D5	Poxvirus D5 protein-like	1	2.2	1	21-33
2090	phoslip	Phospholipase A2	3.4e-49	172.4	1	26-150
2090	RFX_DNA_binding	RFX DNA-binding domain	0.84	2.9	1	55-62
2092	MR_MLE_N	Mandelate racemase / muconate lactoni	1.6e-05	17.0	1	54-157
2092	Peptidase_S26	Signal peptidase I	0.38	3.8	1	99-129
2092	CheR_N	CheR methyltransferase, all-alpha dom	0.4	6.7	1	103-119
2092	MR_MLE	Mandelate racemase / muconate lactoni	2.5e-08	29.9	1	236-298
2094	PP2C	Protein phosphatase 2C	1.2e-71	248.2	1	136-412
2095	EGF	EGF-like domain	0.64	6.1	1	3-29
2095	EGF	EGF-like domain	6.2e-05	20.5	2	35-68

559
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2095	EGF	EGF-like domain	0.00015	19.1	3	94-131
2096	7tm_1	7 transmembrane receptor (rhodopsin f	8.6e-47	138.9	1	83-332
2097	7tm_1	7 transmembrane receptor (rhodopsin f	8.6e-47	138.9	1	83-332
2098	DSL	Delta serrate ligand	0.018	9.3	1	22-37
2098	EGF	EGF-like domain	0.0067	13.2	1	44-71
2098	TIL	Trypsin Inhibitor like cysteine rich	0.33	4.8	1	46-66
2098	DSL	Delta serrate ligand	0.48	4.7	2	56-71
2099	TEP1_N	TEP1 N-terminal domain	0.85	4.7	1	36-65
2099	Pox_A46	Poxvirus A46 family	0.55	2.5	1	61-75
2099	ExoD	Exopolysaccharide synthesis, ExoD	0.82	2.4	1	124-147
2099	RhoGAP	RhoGAP domain	4e-28	95.9	1	161-255
2102	myosin_head	Myosin head (motor domain)	6.3e-56	189.4	1	9-183
2102	ATP_bind2	P-loop ATPase protein family	0.16	4.9	1	75-88
2102	PRK	Phosphoribulokinase / Uridine kinase fa	0.14	5.2	1	77-88
2103	myosin_head	Myosin head (motor domain)	6.3e-56	189.4	1	9-183
2103	ATP_bind2	P-loop ATPase protein family	0.16	4.9	1	75-88
2103	PRK	Phosphoribulokinase / Uridine kinase fa	0.14	5.2	1	77-88
2105	kazal	Kazal-type serine protease inhibitor	8.4e-08	33.5	1	73-117
2105	thyroglobulin_1	Thyroglobulin type-1 repeat	7.7e-19	72.8	1	255-317
2108	BEX	Brain expressed X-linked like family	9.8e-86	266.4	1	79-190
2108	ChaC	ChaC-like protein	0.2	4.5	1	132-157
2108	IlvC	Acetohydroxy acid isomeroreductase, ca	0.14	5.9	1	133-162
2109	LRRCT	Leucine rich repeat C-terminal domain	8.5e-09	28.1	1	45-91
2109	UPF0118	Domain of unknown function DUF20	1	2.9	1	219-242
2112	Inh	Protease inhibitor Inh	0.026	9.0	1	19-44
2112	ank	Ankyrin repeat	0.0042	14.2	1	26-45
2113	DUF370	Domain of unknown function (DUF370)	1	3.5	1	24-39
2113	ApoL	Apolipoprotein L	4e-191	645.1	1	46-348
2113	HupH_C	HupH hydrogenase expression protein,	0.99	2.7	1	119-134
2114	DUF370	Domain of unknown function (DUF370)	1	3.5	1	24-39
2114	ApoL	Apolipoprotein L	4e-191	645.1	1	46-348
2114	HupH_C	HupH hydrogenase expression protein,	0.99	2.7	1	119-134
2115	MAM	MAM domain	1.5e-43	154.8	1	3-102
2116	MAM	MAM domain	1.5e-43	154.8	1	3-102
2117	CBF	CBF/Mak21 family	0.00014	14.4	1	32-65
2118	PLA2_B	Lysophospholipase catalytic domain	7.6e-30	104.2	1	14-143
2118	DUF188	Uncharacterized BCR, Yail/YqxD family CO	0.9	2.9	1	140-151
2119	PLA2_B	Lysophospholipase catalytic domain	7.6e-30	104.2	1	14-143
2119	DUF188	Uncharacterized BCR, Yail/YqxD family CO	0.9	2.9	1	140-151
2121	p450	Cytochrome P450	1.6e-05	16.5	1	31-143
2121	Phage_attach	Phage Head-Tail Attachment	0.97	1.6	1	100-111
2122	ig	Immunoglobulin domain	1.5e-12	49.8	1	38-96
2122	ig	Immunoglobulin domain	2.3e-06	26.7	2	134-213
2122	CD36	CD36 family	0.38	3.9	1	246-271
2122	Neur_chan_mem b	Neurotransmitter-gated ion-channel tra	0.69	2.3	1	261-270
2123	ig	Immunoglobulin domain	1.5e-12	49.8	1	38-96

560

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2123	ig	Immunoglobulin domain	2.3e-06	26.7	2	134-213
2123	CD36	CD36 family	0.38	3.9	1	246-271
2123	Neur_chan_mem b	Neurotransmitter-gated ion-channel tra	0.69	2.3	1	261-270
2124	ig	Immunoglobulin domain	1.5e-12	49.8	1	38-96
2124	ig	Immunoglobulin domain	2.3e-06	26.7	2	134-213
2124	CD36	CD36 family	0.38	3.9	1	246-271
2124	Neur_chan_mem b	Neurotransmitter-gated ion-channel tra	0.69	2.3	1	261-270
2125	C2	C2 domain	0.15	6.6	1	33-48
2125	C2	C2 domain	8.3e-37	125.8	2	92-180
2126	DUF1058	Protein of unknown function (DUF1058)	0.49	2.3	1	80-93
2126	Pep_M12B_prop ep	Reprolysin family propeptide	1.9e-05	17.5	1	155-208
2127	Ifi-6-16	Interferon-induced 6-16 family	3.7e-46	159.7	1	41-123
2127	GLTT	GLTT repeat (6 copies)	0.18	7.7	1	50-78
2127	CRCB	CrcB-like protein	0.18	7.1	1	106-124
2128	abhydrolase	alpha/beta hydrolase fold	0.02	9.2	1	74-127
2128	lipase	Lipase	0.64	3.7	1	98-126
2128	abhydrolase	alpha/beta hydrolase fold	0.0083	10.5	2	167-237
2128	DLH	Dienelactone hydrolase family	0.4	3.6	1	169-196
2128	LIP	Secretory lipase	0.012	8.6	1	178-203
2128	UPF0227	Uncharacterised protein family (UPF02	0.38	4.9	1	179-209
2128	abhydrolase_2	Phospholipase/Carboxylesterase	0.015	10.1	1	180-203
2128	Peptidase_M10_N	Matrix metalloprotease, N-terminal do	0.63	2.5	1	209-230
2129	abhydrolase	alpha/beta hydrolase fold	0.02	9.2	1	74-127
2129	lipase	Lipase	0.64	3.7	1	98-126
2129	abhydrolase	alpha/beta hydrolase fold	0.0083	10.5	2	167-237
2129	DLH	Dienelactone hydrolase family	0.4	3.6	1	169-196
2129	LIP	Secretory lipase	0.012	8.6	1	178-203
2129	UPF0227	Uncharacterised protein family (UPF02	0.38	4.9	1	179-209
2129	abhydrolase_2	Phospholipase/Carboxylesterase	0.015	10.1	1	180-203
2129	Peptidase_M10_N	Matrix metalloprotease, N-terminal do	0.63	2.5	1	209-230
2130	Collagen	Collagen triple helix repeat (20 copie	1.4e-06	27.0	1	1-38
2130	Collagen	Collagen triple helix repeat (20 copie	2.5e-05	22.3	2	39-74
2130	SRCR	Scavenger receptor cysteine-rich domai	2.6e-16	59.1	1	90-126
2131	Collagen	Collagen triple helix repeat (20 copie	1.4e-06	27.0	1	1-38
2131	Collagen	Collagen triple helix repeat (20 copie	2.5e-05	22.3	2	39-74
2131	SRCR	Scavenger receptor cysteine-rich domai	2.6e-16	59.1	1	90-126
2132	RICH	RICH domain	0.3	5.4	1	290-320
2132	DUF260	Protein of unknown function DUF260	0.047	7.1	1	425-447
2132	Ter	DNA replication terminus site-binding	0.019	7.5	1	427-450
2132	Tropomyosin	Tropomyosin	0.27	4.7	1	468-506
2132	Adeno_PIX	Adenovirus hexon-associated protein (0.044	8.0	1	482-506
2132	AgrD	Staphylococcal AgrD protein	0.83	5.2	1	501-508
2132	K-box	K-box region	0.0023	12.6	1	569-602
2132	Tfb2	Transcription factor Tfb2	0.98	-1.2	1	591-610
2132	RRF	Ribosome recycling factor	0.5	5.0	2	696-727
2132	G-gamma	GGL domain	0.33	5.0	1	717-738
2132	DUF260	Protein of unknown function DUF260	0.39	4.2	2	821-843
2132	bZIP	bZIP transcription factor	0.52	5.5	2	835-873

561
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2132	Lipoprotein_11	Lepidopteran low molecular weight (30	1	2.9	1	850-867
2132	DNA_ligase_N	NAD-dependent DNA ligase adenylation	0.081	5.7	1	868-888
2133	KRAB	KRAB box	2.9e-27	100.7	1	61-101
2133	Androgen_recep	Androgen receptor	0.71	0.7	1	70-80
2133	TFIIS	Transcription factor S-II (TFIIS)	0.73	5.1	1	324-334
2133	zf-C2H2	Zinc finger, C2H2 type	3.5e-05	25.4	1	324-346
2133	zf-C2H2	Zinc finger, C2H2 type	1.3e-06	31.2	2	352-374
2133	zf-BED	BED zinc finger	0.33	5.7	1	354-375
2133	mRNA_cap_enzyme	mRNA capping enzyme, catalytic domain	0.56	0.5	1	377-392
2133	XPA_N	XPA protein N-terminal	0.78	5.1	2	377-389
2133	zf-C2H2	Zinc finger, C2H2 type	2.9e-07	33.8	3	380-402
2133	TFIIS	Transcription factor S-II (TFIIS)	0.89	4.8	3	408-418
2133	zf-C2H2	Zinc finger, C2H2 type	2e-06	30.4	4	408-430
2133	zf-C2H2	Zinc finger, C2H2 type	1.6e-05	26.8	5	436-458
2133	mRNA_cap_enzyme	mRNA capping enzyme, catalytic domain	0.56	0.5	2	461-476
2133	XPA_N	XPA protein N-terminal	0.78	5.1	4	461-473
2133	zf-C2H2	Zinc finger, C2H2 type	5.4e-07	32.7	6	464-486
2133	TFIIS	Transcription factor S-II (TFIIS)	0.29	6.5	5	492-502
2133	zf-C2H2	Zinc finger, C2H2 type	1.1e-06	31.5	7	492-514
2133	XPA_N	XPA protein N-terminal	0.13	7.8	6	517-529
2133	TFIIS	Transcription factor S-II (TFIIS)	0.57	5.5	6	520-530
2133	zf-C2H2	Zinc finger, C2H2 type	9.2e-07	31.8	8	520-542
2133	XPA_N	XPA protein N-terminal	0.97	4.8	7	545-557
2133	TFIIS	Transcription factor S-II (TFIIS)	0.14	7.6	7	548-558
2133	zf-C2H2	Zinc finger, C2H2 type	4.4e-06	29.1	9	548-570
2133	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.38	3.3	1	560-581
2133	zf-C2H2	Zinc finger, C2H2 type	1.1e-06	31.5	10	576-598
2133	TFIIS	Transcription factor S-II (TFIIS)	0.054	9.0	8	604-614
2133	zf-C2H2	Zinc finger, C2H2 type	2.9e-07	33.8	11	604-626
2133	zf-BED	BED zinc finger	0.64	4.8	3	609-627
2133	DC1	1/2 604 619 .. 19 44	0.16	6.2	2	632-647
2133	zf-C2H2	Zinc finger, C2H2 type	0.00082	19.9	12	632-655
2137	aminotran_3	Aminotransferase class-III	1.2e-09	31.3	1	55-114
2137	OATP_N	Organic Anion Transporter Polypeptide	0.81	4.0	1	140-158
2137	aminotran_3	Aminotransferase class-III	8.1e-63	208.6	2	181-409
2138	aminotran_3	Aminotransferase class-III	1.2e-09	31.3	1	55-114
2138	OATP_N	Organic Anion Transporter Polypeptide	0.81	4.0	1	140-158
2138	aminotran_3	Aminotransferase class-III	8.1e-63	208.6	2	181-409
2139	trypsin	Trypsin	1.3e-25	79.1	1	8-114
2140	Glycos_transf_1	Glycosyl transferases group 1	1.7e-17	64.4	1	99-194
2141	MHYT	Bacterial signalling protein N termina	0.6	4.2	1	291-328
2142	EGF	EGF-like domain	8.8e-09	34.4	1	1-30
2142	EGF	EGF-like domain	1.5e-07	30.0	2	41-72
2142	EGF	EGF-like domain	0.0091	12.7	3	82-107
2142	EB	EB module	0.077	7.1	2	116-148
2142	EGF	EGF-like domain	1.3e-07	30.2	4	116-148
2142	EGF	EGF-like domain	0.022	11.3	5	157-181
2143	AdoHcyase	S-adenosyl-L-homocysteine hydrolase	0.0022	9.6	1	1-15
2143	AdoHcyase_NA D	S-adenosyl-L-homocysteine hydrolase, NA	0.0012	13.8	1	16-27
2144	UQ_con	Ubiquitin-conjugating enzyme	0.0058	11.9	1	31-61

562
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2144	UQ_con	Ubiquitin-conjugating enzyme	5.8e-25	91.7	2	89-156
2145	Prominin	Prominin	2.1e-113	364.0	1	25-213
2145	SPDY	Domain of unknown function (DUF317)	0.15	6.5	1	87-100
2145	Prominin	Prominin	6.7e-30	94.4	2	214-286
2145	Prominin	Prominin	1.9e-139	448.0	3	287-510
2146	fibrinogen_C	Fibrinogen beta and gamma chains, C-ter	6.5e-54	184.0	1	13-231
2147	fibrinogen_C	Fibrinogen beta and gamma chains, C-ter	6.5e-54	184.0	1	13-231
2148	fibrinogen_C	Fibrinogen beta and gamma chains, C-ter	6.5e-54	184.0	1	13-231
2150	DUF381	Domain of unknown function (DUF381)	0.48	4.4	1	29-35
2151	aa_permeases	Amino acid permease	7e-24	89.2	1	6-294
2151	Pox_I5	Poxvirus protein I5	0.24	6.0	1	85-102
2151	serine_carbpept	Serine carboxypeptidase	0.41	2.3	1	301-321
2153	spectrin	Spectrin repeat	0.4	5.5	1	410-463
2154	spectrin	Spectrin repeat	0.4	5.5	1	410-463
2155	Peptidase_M20	Peptidase family M20/M25/M40	0.00038	14.5	1	39-120
2156	sugar_tr	Sugar (and other) transporter	0.11	5.5	1	47-103
2156	Octopine_DH	NAD/NADP octopine/nopaline dehydrogenas	0.26	4.6	1	153-169
2156	sugar_tr	Sugar (and other) transporter	5e-08	28.1	2	201-336
2159	bromodomain	Bromodomain	9.5e-45	158.8	1	74-163
2159	bromodomain	Bromodomain	3e-40	143.5	2	367-456
2159	Alpha_adaptin_C	Alpha adaptin AP2, C-terminal domain	0.48	2.6	1	406-418
2159	Phage_X	Phage X family	0.97	3.7	1	449-480
2159	eIF3c_N	Eukaryotic translation initiation fac	0.51	1.2	1	484-570
2159	Vitellogenin_N	Lipoprotein amino terminal region	0.61	1.5	1	495-550
2159	Herpes_U44	Herpes virus U44 protein	0.47	3.1	1	526-540
2161	ig	Immunoglobulin domain	6.4e-06	25.0	1	58-118
2164	pkinase	Protein kinase domain	2.6e-38	136.6	1	7-108
2164	TMP	TMP repeat	0.37	8.0	1	74-84
2165	pkinase	Protein kinase domain	2.6e-38	136.6	1	7-108
2165	TMP	TMP repeat	0.37	8.0	1	74-84
2166	glutaredoxin	Glutaredoxin	0.00075	15.0	1	12-65
2166	GST_N	Glutathione S-transferase, N-terminal	0.019	11.1	1	13-63
2166	GST_C	Glutathione S-transferase, C-terminal	0.00013	17.6	1	189-281
2166	UL21	Herpesvirus UL21	0.98	0.3	1	212-240
2167	PadR	Transcriptional regulator PadR-like f	0.22	6.1	1	18-31
2167	Collagen	Collagen triple helix repeat (20 copi	2.4e-05	22.3	1	43-76
2167	Collagen	Collagen triple helix repeat (20 copi	1.5e-07	30.6	2	77-122
2167	C1q	C1q domain	2.9e-72	250.2	1	132-257
2167	TOBE	TOBE domain	0.5	6.3	1	223-242
2169	Sec6	Exocyst complex component Sec6	0.71	2.3	1	166-194
2169	BRCT	BRCA1 C Terminus (BRCT) domain	0.0053	11.4	1	278-315
2169	Chitin_bind_3	Chitin binding domain	0.95	2.1	1	308-321
2169	BRCT	BRCA1 C Terminus (BRCT) domain	0.00072	14.3	2	329-369
2169	BRCT	BRCA1 C Terminus (BRCT) domain	5.7e-19	65.1	3	378-451
2169	BRCT	BRCA1 C Terminus (BRCT) domain	4e-19	65.6	4	536-622
2169	RinB	Transcriptional activator RinB	0.33	5.4	1	595-646

563
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2169	BRCT	BRCA1 C Terminus (BRCT) domain	0.028	9.0	5	645-680
2170	Sec6	Exocyst complex component Sec6	0.71	2.3	1	166-194
2170	BRCT	BRCA1 C Terminus (BRCT) domain	0.0053	11.4	1	278-315
2170	Chitin_bind_3	Chitin binding domain	0.95	2.1	1	308-321
2170	BRCT	BRCA1 C Terminus (BRCT) domain	0.00072	14.3	2	329-369
2170	BRCT	BRCA1 C Terminus (BRCT) domain	5.7e-19	65.1	3	378-451
2170	BRCT	BRCA1 C Terminus (BRCT) domain	4e-19	65.6	4	536-622
2170	RinB	Transcriptional activator RinB	0.33	5.4	1	595-646
2170	BRCT	BRCA1 C Terminus (BRCT) domain	0.028	9.0	5	645-680
2172	LRRCT	Leucine rich repeat C-terminal domain	8.5e-09	28.1	1	45-91
2172	UPF0118	Domain of unknown function DUF20	1	2.9	1	219-242
2173	ig	Immunoglobulin domain	7.9e-06	24.7	1	39-93
2173	Na_Ca_Ex	Sodium/calcium exchanger protein	0.86	4.3	1	133-148
2173	COX17	Cytochrome C oxidase copper chaperone	0.68	3.6	1	196-209
2174	TB2_DP1_HVA22	TB2/DP1, HVA22 family	3.8e-34	123.6	1	18-111
2174	ELM2	ELM2 domain	0.53	5.2	1	114-139
2175	An_peroxidase	Animal haem peroxidase	1.3e-91	311.6	1	2-232
2175	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	24-32
2175	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	117-134
2176	An_peroxidase	Animal haem peroxidase	1.3e-91	311.6	1	2-232
2176	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	24-32
2176	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	117-134
2177	An_peroxidase	Animal haem peroxidase	1.3e-91	311.6	1	2-232
2177	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	24-32
2177	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	117-134
2178	DUF846	Eukaryotic protein of unknown functio	0.0084	8.0	1	55-84
2179	UPF0137	Uncharacterised protein family (UPF01	0.04	7.4	1	341-366
2179	PS_pyruv_trans	Polysaccharide pyruvyl transferase	0.55	3.3	1	355-411
2180	COX17	Cytochrome C oxidase copper chaperone	0.51	4.0	1	39-60
2180	RIIa	Regulatory subunit of type II PKA R-s	1e-14	54.8	1	67-104
2180	SURF6	Surfeit locus protein 6	0.027	7.2	1	84-155
2180	cNMP_binding	Cyclic nucleotide-binding domain	7.2e-31	112.5	1	194-282
2180	RNA_pol_Rpb2_4	RNA polymerase Rpb2, domain 4	0.28	6.2	1	226-233
2180	cNMP_binding	Cyclic nucleotide-binding domain	9.4e-32	115.7	2	312-406
2180	Methyltransf_1	6-O-methylguanine DNA methyltransfera	0.64	4.3	1	367-379
2181	PDZ	PDZ domain (Also known as DHR or GLGF)	6.7e-12	43.7	1	5-86
2182	PLAT	PLAT/LH2 domain	1.7e-31	108.4	1	2-111
2182	lipxygenase	Lipxygenase	3.9e-194	655.1	1	113-624
2182	DUF181	Uncharacterized ACR, COG1944	0.81	2.4	1	221-232
2182	PG_binding_1	Putative peptidoglycan binding domain	0.5	5.6	1	395-411
2183	PLAT	PLAT/LH2 domain	1.7e-31	108.4	1	2-111
2183	lipxygenase	Lipxygenase	3.9e-194	655.1	1	113-624
2183	DUF181	Uncharacterized ACR, COG1944	0.81	2.4	1	221-232
2183	PG_binding_1	Putative peptidoglycan binding domain	0.5	5.6	1	395-411
2184	PLAT	PLAT/LH2 domain	1.7e-31	108.4	1	2-111
2184	lipxygenase	Lipxygenase	3.9e-	655.1	1	113-624

564

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
			194			
2184	DUF181	Uncharacterized ACR, COG1944	0.81	2.4	1	221-232
2184	PG_binding_1	Putative peptidoglycan binding domain	0.5	5.6	1	395-411
2186	TFIIS	Transcription factor S-II (TFIIS)	1	4.6	1	11-21
2186	DUF536	Protein of unknown function, DUF536	0.19	7.9	1	220-257
2186	FCH	Fes/CIP4 homology domain	0.5	5.6	1	265-284
2192	Aa_trans	Transmembrane amino acid transporter prote	3.8e-09	33.4	1	4-56
2193	EGF	EGF-like domain	0.024	11.2	1	42-57
2193	EGF	EGF-like domain	1.3e-06	26.6	2	60-88
2193	EGF	EGF-like domain	1.2e-09	37.5	3	95-128
2193	Cripto	Cripto growth factor	0.86	3.4	1	101-132
2193	laminin_EGF	1/3 32 60 .. 2 43	0.025	9.9	2	106-130
2193	EGF	EGF-like domain	5.5e-07	27.9	4	135-171
2194	M	M protein repeat	0.8	7.1	1	64-84
2194	PP1_inhibitor	PKC-activated protein phosphatase-1 i	0.78	2.2	1	303-319
2194	bZIP	1/2 65 82 .. 48 65	0.32	6.2	2	397-415
2194	TSC22	TSC-22/dip/bun family	0.045	7.2	1	398-415
2195	ank	Ankyrin repeat	0.0017	15.6	2	206-231
2195	G-patch	G-patch domain	2e-16	58.7	1	319-363
2195	Anti-silence	Anti-silencing protein, ASF1-like	0.18	5.1	1	365-378
2196	endotoxin	delta endotoxin	0.85	2.3	1	134-151
2197	Peptidase_M24	metallopeptidase family M24	5.5e-69	239.3	1	103-342
2197	DUF120	Domain of unknown function DUF120	0.089	7.1	1	184-195
2199	PAAD_DAPIN	PAAD/DAPIN/Pyrin domain	1.3e-11	41.6	1	18-103
2199	DHHA1	DHHA1 domain	0.61	5.4	1	67-87
2199	UPF0160	Uncharacterised protein family (UPF01	1	2.3	1	75-86
2199	RNA_helicase	RNA helicase	0.03	7.9	1	195-215
2199	NACHT	NACHT domain	3.8e-74	252.4	1	196-365
2199	AAA	ATPase family associated with various	0.15	5.2	1	197-215
2199	Peptidase_S15	X-Pro dipeptidyl-peptidase (S15 famil	0.64	2.1	1	929-984
2200	PDZ	PDZ domain (Also known as DHR or GLGF	8.1e-22	78.5	1	35-114
2200	CDC50	LEM3 (ligand-effect modulator 3) fami	1	2.1	1	101-116
2200	DUF100	Protein of unknown function DUF100	0.2	4.1	1	117-130
2201	DIE2_ALG10	DIE2/ALG10 family	1.5e-54	191.4	1	62-142
2201	DUF718	Protein of unknown function (DUF718)	0.64	4.4	1	70-77
2202	rrm	RNA recognition motif. (a.k.a. RRM, R	1.3e-09	36.2	1	76-143
2202	RbsD_FucU	RbsD / FucU transport protein family	0.53	3.4	1	138-162
2202	HemX	HemX	0.37	3.5	1	157-188
2202	rrm	RNA recognition motif. (a.k.a. RRM, R	4.6e-13	48.6	2	201-268
2202	rrm	RNA recognition motif. (a.k.a. RRM, R	4.3e-13	48.7	3	354-421
2202	rrm	RNA recognition motif. (a.k.a. RRM, R	1.4e-06	25.5	4	471-539
2203	C_tripleX	Cysteine rich repeat	2e-05	17.8	1	76-93
2203	Bowman-Birk_leg	Bowman-Birk serine protease inhibitor	1	4.0	1	85-100
2203	laminin_EGF	Laminin EGF-like (Domains III and V)	0.32	6.1	1	97-110
2203	EGF	EGF-like domain	8.7e-06	23.6	2	115-143
2203	TIL	Trypsin Inhibitor like cysteine rich	0.0035	11.0	1	134-155
2203	EGF	EGF-like domain	7.5e-05	20.2	3	155-189
2203	TIL	Trypsin Inhibitor like cysteine rich	0.26	5.1	2	168-195
2203	toxin_5	Scorpion short toxin	0.34	4.4	1	170-175
2203	EGF	EGF-like domain	4.4e-05	21.1	4	195-228
2203	EGF	EGF-like domain	9.7e-09	34.3	5	240-275

565
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2203	MAM	MAM domain	9.2e-38	135.6	1	421-566
2204	C_tripleX	Cysteine rich repeat	2e-05	17.8	1	76-93
2204	Bowman-Birk_leg	Bowman-Birk serine protease inhibitor	1	4.0	1	85-100
2204	laminin_EGF	Laminin EGF-like (Domains III and V)	0.32	6.1	1	97-110
2204	EGF	EGF-like domain	8.7e-06	23.6	2	115-143
2204	TIL	Trypsin Inhibitor like cysteine rich	0.0035	11.0	1	134-155
2204	EGF	EGF-like domain	7.5e-05	20.2	3	155-189
2204	TIL	Trypsin Inhibitor like cysteine rich	0.26	5.1	2	168-195
2204	toxin_5	Scorpion short toxin	0.34	4.4	1	170-175
2204	EGF	EGF-like domain	4.4e-05	21.1	4	195-228
2204	EGF	EGF-like domain	9.7e-09	34.3	5	240-275
2204	MAM	MAM domain	9.2e-38	135.6	1	421-566
2205	TH1	TH1 protein	0.91	0.2	1	315-328
2205	Neg_reg	Negative transcriptional regulator	1	2.3	1	587-596
2205	zf-MYND	MYND finger	2e-08	27.7	1	654-688
2206	TH1	TH1 protein	0.91	0.2	1	315-328
2206	Neg_reg	Negative transcriptional regulator	1	2.3	1	587-596
2206	zf-MYND	MYND finger	2e-08	27.7	1	654-688
2207	TH1	TH1 protein	0.91	0.2	1	315-328
2207	Neg_reg	Negative transcriptional regulator	1	2.3	1	587-596
2207	zf-MYND	MYND finger	2e-08	27.7	1	654-688
2208	TH1	TH1 protein	0.91	0.2	1	315-328
2208	Neg_reg	Negative transcriptional regulator	1	2.3	1	587-596
2208	zf-MYND	MYND finger	2e-08	27.7	1	654-688
2209	Urotensin_II	Urotensin II	0.36	5.4	1	82-92
2209	fn2	Fibronectin type II domain	0.55	3.5	1	83-91
2210	RNA_helicase	RNA helicase	0.03	7.9	1	91-111
2210	NACHT	NACHT domain	3.8e-74	252.4	1	92-261
2210	AAA	ATPase family associated with various	0.15	5.2	1	93-111
2211	disintegrin	Disintegrin	3.3e-36	123.1	1	4-79
2211	EGF	EGF-like domain	0.0023	14.8	1	231-258
2213	zf-MYND	MYND finger	1	3.8	1	38-45
2213	ank	Ankyrin repeat	4.4e-06	24.9	1	159-187
2213	ank	Ankyrin repeat	6.9e-09	35.0	2	191-223
2213	ank	Ankyrin repeat	0.15	8.6	3	224-256
2213	ank	Ankyrin repeat	9.7e-10	38.0	4	258-290
2213	ank	Ankyrin repeat	0.00014	19.5	5	291-336
2213	LoIA	Outer membrane lipoprotein carrier pr	1	3.0	1	317-340
2213	ank	Ankyrin repeat	3.8e-08	32.3	6	337-369
2213	ank	Ankyrin repeat	0.49	6.8	7	370-402
2214	interferon	Interferon alpha/beta domain	3.7e-42	145.6	1	27-116
2215	DUF602	Protein of unknown function, DUF602	1.3e-202	683.2	1	15-303
2215	Bromo_MP	Bromovirus movement protein	0.062	6.4	1	21-47
2216	DUF846	Eukaryotic protein of unknown function	0.012	7.5	1	120-150
2218	acid_phosphat	Histidine acid phosphatase	5.5e-13	45.0	1	137-232
2219	PH	PH domain	1.9e-20	68.8	1	78-238
2219	ArfGap	Putative GTPase activating protein fo	5.4e-50	174.5	1	259-379
2219	ank	Ankyrin repeat	1.9e-09	36.9	1	418-450
2219	ank	Ankyrin repeat	0.022	11.6	2	451-475
2219	SapB_2	Saposin-like type B, region 2	0.33	6.5	1	464-475
2220	PH	PH domain	1.9e-20	68.8	1	78-238

566
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2220	ArfGap	Putative GTPase activating protein fo	5.4e-50	174.5	1	259-379
2220	ank	Ankyrin repeat	1.9e-09	36.9	1	418-450
2220	ank	Ankyrin repeat	0.022	11.6	2	451-475
2220	SapB_2	Saposin-like type B, region 2	0.33	6.5	1	464-475
2221	PH	PH domain	1.9e-20	68.8	1	78-238
2221	ArfGap	Putative GTPase activating protein fo	5.4e-50	174.5	1	259-379
2221	ank	Ankyrin repeat	1.9e-09	36.9	1	418-450
2221	ank	Ankyrin repeat	0.022	11.6	2	451-475
2221	SapB_2	Saposin-like type B, region 2	0.33	6.5	1	464-475
2222	PH	PH domain	1.9e-20	68.8	1	78-238
2222	ArfGap	Putative GTPase activating protein fo	5.4e-50	174.5	1	259-379
2222	ank	Ankyrin repeat	1.9e-09	36.9	1	418-450
2222	ank	Ankyrin repeat	0.022	11.6	2	451-475
2222	SapB_2	Saposin-like type B, region 2	0.33	6.5	1	464-475
2223	Reprolysin	Reprolysin (M12B) family zinc metallo	2.4e-35	127.6	1	3-83
2223	Astacin	Astacin (Peptidase family M12A)	0.21	5.0	1	23-37
2223	Phi_1	Phosphate-induced protein 1 conserved	0.51	3.3	1	71-83
2223	disintegrin	Disintegrin	0.0019	12.9	1	101-136
2224	Uteroglobin	Uteroglobin family	6.6e-09	29.8	1	1-88
2225	C1q	C1q domain	6.3e-06	23.8	1	98-138
2226	Ornatin	Ornatin	0.55	4.8	1	99-106
2227	Ornatin	Ornatin	0.55	4.8	1	99-106
2228	Gag_MA	Matrix protein (MA), p15	0.11	6.5	1	96-152
2229	Seryl_tRNA_N	Seryl-tRNA synthetase N-terminal doma	0.92	5.7	1	241-258
2229	pentaxin	Pentaxin family	4.3e-24	83.3	1	363-526
2229	Avirulence	Xanthomonas avirulence protein, Avr/P	0.07	3.6	1	501-515
2233	ion_trans	Ion transport protein	0.001	14.0	1	22-141
2233	Sarcoplin	Sarcoplin	0.56	5.3	1	95-123
2234	ion_trans	Ion transport protein	0.001	14.0	1	22-141
2234	Sarcoplin	Sarcoplin	0.56	5.3	1	95-123
2235	zf-C2H2	Zinc finger, C2H2 type	0.033	13.4	2	100-123
2235	TFIID-31	Transcription initiation factor IID, 3	0.28	5.7	1	120-135
2235	zf-C2H2	Zinc finger, C2H2 type	0.14	10.9	3	134-156
2238	asp	Eukaryotic aspartyl protease	1.1e-24	87.5	1	1-67
2239	Sulfatase	Sulfatase	4.5e-05	18.1	1	57-122
2240	Zn_carbOpept	Zinc carboxypeptidase	3.7e-57	193.6	1	13-156
2241	Zn_carbOpept	Zinc carboxypeptidase	3.7e-57	193.6	1	13-156
2242	NifU_N	NifU-like N terminal domain	1.7e-80	277.6	1	34-160
2244	zf-C2H2	Zinc finger, C2H2 type	0.00035	21.4	1	56-81
2244	zf-C2H2	Zinc finger, C2H2 type	0.012	15.2	2	90-117
2244	zf-C2H2	Zinc finger, C2H2 type	0.0039	17.1	3	123-147
2245	pkinase	Protein kinase domain	3.2e-90	309.9	1	49-341
2245	Glyco_hydro_15	Glycosyl hydrolases family 15	0.18	4.4	1	501-551
2248	C1q	C1q domain	5.1e-23	86.7	1	27-135
2249	Allantoicase	Allantoicase repeat	0.014	9.0	1	13-23
2249	Allantoicase	Allantoicase repeat	1.3e-57	196.4	2	46-206
2250	DNA_ligase_A_C	ATP dependent DNA ligase C terminal r	0.67	5.4	1	25-48
2250	ig	Immunoglobulin domain	0.00019	19.5	1	51-165
2250	ig	Immunoglobulin domain	0.15	8.7	2	196-257
2250	ig	Immunoglobulin domain	0.0031	15.0	3	289-349
2250	SK_channel	Calcium-activated SK potassium channe	0.035	7.1	1	377-397

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2251	PH	PH domain	2.4e-24	81.6	1	43-153
2251	HS2ST	Heparan sulfate 2-O-sulfotransferase	0.27	4.4	1	160-182
2251	LMP	LMP repeated region	0.0012	14.2	1	180-201
2251	DUF603	Protein of unknown function, DUF603	0.04	6.4	1	193-207
2251	Pox_A_type_inc	Viral A-type inclusion protein repeat	0.32	7.2	1	193-207
2251	IQ	IQ calmodulin-binding motif	5e-05	20.1	1	226-246
2251	RhoGEF	RhoGEF domain	1.2e-69	236.9	1	267-448
2251	DUF674	Protein of unknown function (DUF674)	0.82	1.4	1	295-305
2251	Stig1	Stigma-specific protein, Stig1	0.6	2.3	1	396-441
2251	PH	PH domain	2.3e-13	45.3	2	480-608
2251	RasGEFN	Guanine nucleotide exchange factor fo	1.1e-19	71.3	1	653-708
2251	RasGEF	RasGEF domain	7.2e-89	305.4	1	1019-1204
2251	Adeno_terminal	Adenoviral DNA terminal protein	1	1.7	1	1195-1227
2252	DUF630	Protein of unknown function (DUF630)	0.7	4.3	1	584-597
2252	FGF	Fibroblast growth factor	0.37	4.4	1	620-635
2252	tRNA-synt_2	tRNA synthetases class II (D, K and N	0.74	3.5	1	646-658
2252	Omega-atracotox	Omega-atracotoxin	0.15	5.1	1	751-758
2253	K_tetra	K ⁺ channel tetramerisation domain	2e-34	121.3	1	26-114
2253	BTB	BTB/POZ domain	0.0015	14.2	1	74-125
2254	PXA	PXA domain	0.01	10.2	1	90-110
2254	Vps52	Vps52 / Sac2 family	0	1089.2	1	100-609
2254	trp_syntA	Tryptophan synthase alpha chain	0.78	3.1	1	179-216
2254	DUF965	Bacterial protein of unknown function	0.33	4.5	1	291-304
2255	NosL	NosL	0.29	4.9	1	104-128
2255	NAC	NAC domain	0.76	5.5	1	150-172
2255	DUF240	MG032/MG096/MG288 family 2	0.17	6.7	1	176-191
2256	NosL	NosL	0.29	4.9	1	104-128
2256	NAC	NAC domain	0.76	5.5	1	150-172
2256	DUF240	MG032/MG096/MG288 family 2	0.17	6.7	1	176-191
2258	ZZ	Zinc finger, ZZ type	1e-12	48.2	1	5-50
2258	SoxD	Sarcosine oxidase, delta subunit fami	0.97	4.2	1	79-86
2258	zf-C2H2	Zinc finger, C2H2 type	0.00067	20.3	1	80-103
2258	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.3	3.6	1	95-115
2258	SPDY	Domain of unknown function (DUF317)	0.6	4.4	1	119-133
2258	Di19	Drought induced 19 protein (Di19)	0.00056	13.0	1	314-330
2261	RmuC	RmuC family	0.79	3.1	1	16-46
2261	IBN_NT	Importin-beta N-terminal domain	2.1e-27	99.5	1	34-113
2261	Peripla_BP_like	Periplasmic binding proteins and suga	0.21	4.7	1	142-173
2262	Las1	Las1-like	1.6e-94	320.7	1	55-203
2262	MuDR	MuDR family transposase	0.17	5.5	1	231-263
2262	BAR	BAR domain	0.21	5.2	1	347-363
2262	Adeno_E1B_19K	Adenovirus E1B 19K protein / small t-	0.43	4.6	1	534-558
2262	META	Domain of unknown function (306)	0.91	5.3	1	632-663
2263	ank	Ankyrin repeat	0.00019	19.0	1	1-33
2263	DMRL_synthase	6,7-dimethyl-8-ribityllumazine synthas	0.35	5.0	1	33-48
2263	hormone	Somatotropin hormone family	0.23	2.6	1	85-115
2265	ig	Immunoglobulin domain	4.1e-05	22.0	1	10-78
2265	ig	Immunoglobulin domain	3.7e-10	40.9	2	113-172
2265	ig	Immunoglobulin domain	0.0018	15.9	3	211-272
2265	ig	Immunoglobulin domain	3.7e-08	33.4	4	309-370

568

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2265	DNA_pol_B_2	DNA polymerase type B, organellar and	0.018	7.9	1	326-382
2265	OapA	Opacity-associated protein A	0.44	2.4	1	335-357
2265	ig	Immunoglobulin domain	0.0012	16.6	5	404-465
2265	ig	Immunoglobulin domain	7.7e-07	28.5	6	500-564
2269	efhand	EF hand	2.8e-08	33.9	1	60-88
2269	COX17	Cytochrome C oxidase copper chaperone	0.42	4.2	1	85-92
2269	efhand	EF hand	0.0033	15.3	2	96-124
2269	efhand	EF hand	8.5e-05	21.1	3	133-161
2269	PCRF	PCRF domain	0.43	6.1	1	160-176
2269	DUF21	Domain of unknown function DUF21	0.18	6.4	1	165-189
2269	efhand	EF hand	5e-09	36.7	4	169-197
2270	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-14	51.3	1	15-61
2270	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	6.8e-52	182.6	2	95-275
2270	Flavodoxin_2	Flavodoxin-like fold	0.66	3.3	1	369-384
2270	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-05	19.1	3	399-440
2270	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.9e-49	174.5	4	501-654
2270	Flavodoxin_2	Flavodoxin-like fold	0.66	3.3	2	748-763
2270	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-05	19.1	5	778-819
2271	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-14	51.3	1	15-61
2271	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	6.8e-52	182.6	2	95-275
2271	Flavodoxin_2	Flavodoxin-like fold	0.66	3.3	1	369-384
2271	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-05	19.1	3	399-440
2271	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.9e-49	174.5	4	501-654
2271	Flavodoxin_2	Flavodoxin-like fold	0.66	3.3	2	748-763
2271	UPF0061	Uncharacterized ACR, YdiU/UPF0061 fam	1.2e-05	19.1	5	778-819
2272	7tm_1	7 transmembrane receptor (rhodopsin fam	9.7e-25	72.7	1	1-107
2273	LRR	Leucine Rich Repeat	0.00057	16.1	1	40-63
2273	LRR	Leucine Rich Repeat	0.004	13.3	3	88-113
2273	LRR	Leucine Rich Repeat	0.84	5.4	4	114-131
2274	AMP-binding	AMP-binding enzyme	6.9e-18	64.1	1	20-135
2275	cytochrome_c	Cytochrome c	0.92	3.7	1	94-110
2275	cNMP_binding	Cyclic nucleotide-binding domain	1.5e-15	57.4	1	126-216
2275	RasGEFN	Guanine nucleotide exchange factor fo	0.00023	17.5	1	241-285
2275	Pseu_avirulence	Avirulence protein	0.91	1.9	1	272-285
2275	PDZ	PDZ domain (Also known as DHR or GLGF	2e-09	35.0	1	361-412
2276	cytochrome_c	Cytochrome c	0.92	3.7	1	94-110
2276	cNMP_binding	Cyclic nucleotide-binding domain	1.5e-15	57.4	1	126-216
2276	RasGEFN	Guanine nucleotide exchange factor fo	0.00023	17.5	1	241-285
2276	Pseu_avirulence	Avirulence protein	0.91	1.9	1	272-285
2276	PDZ	PDZ domain (Also known as DHR or	2e-09	35.0	1	361-412

569

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
		GLGF				
2280	Ricin_B_lectin	QXW lectin repeat	0.14	8.2	1	50-77
2280	MCR_beta_N	Methyl-coenzyme M reductase beta subun	0.98	2.1	1	68-76
2281	ArsA_ATPase	Anion-transporting ATPase	0.54	3.1	1	26-52
2281	ParA	ParA family ATPase	1.4e-25	89.8	1	111-202
2281	SCF	Stem cell factor	1.1e-27	90.4	1	206-259
2281	FH2	Formin Homology 2 Domain	0.027	8.8	1	221-238
2282	cadherin	Cadherin domain	6e-20	71.3	1	3-81
2282	cadherin	Cadherin domain	5.9e-21	74.8	2	118-191
2283	cadherin	Cadherin domain	6e-20	71.3	1	3-81
2283	cadherin	Cadherin domain	5.9e-21	74.8	2	118-191
2284	PH	PH domain	7.9e-10	33.6	1	4-92
2284	DUF1041	Domain of Unknown Function (DUF1041)	1.6e-07	28.1	1	206-237
2285	Renal_dipeptase	Renal dipeptidase	9.3e-05	15.8	1	74-102
2286	aa_permeases	Amino acid permease	7e-24	89.2	1	6-294
2286	Pox_I5	Poxvirus protein I5	0.24	6.0	1	85-102
2286	serine_carbpept	Serine carboxypeptidase	0.41	2.3	1	301-321
2287	THF_DHG_CYH	Tetrahydrofolate dehydrogenase/cycloh	2.3e-11	32.7	1	62-123
2287	THF_DHG_CYH_C	Tetrahydrofolate dehydrogenase/cycloh	6.1e-10	36.6	1	125-171
2287	FTHFS	Formate--tetrahydrofolate ligase	0	1365.1	1	302-921
2288	acid_phosphat	Histidine acid phosphatase	0.038	6.9	1	391-407
2288	FMN_red	NADPH-dependent FMN reductase	0.94	3.3	1	438-459
2288	acid_phosphat	Histidine acid phosphatase	0.02	7.9	2	525-594
2288	Ribosomal_L6	Ribosomal protein L6	0.21	7.2	1	774-814
2290	PI-PLC-X	Phosphatidylinositol-specific phospholipase	3.8e-14	50.6	1	1-33
2292	ABG_transport	AbgT putative transporter family	0.81	1.2	1	21-34
2292	7tm_1	7 transmembrane receptor (rhodopsin f	1.6e-30	90.1	1	48-297
2292	HECT	HECT-domain (ubiquitin-transferase)	0.15	5.5	1	281-298
2293	tsp_3	Thrombospondin type 3 repeat	0.00058	15.9	1	13-25
2293	tsp_3	Thrombospondin type 3 repeat	0.0033	13.4	2	36-48
2293	tsp_3	Thrombospondin type 3 repeat	0.0011	15.0	3	51-66
2293	tsp_3	Thrombospondin type 3 repeat	0.00057	15.9	4	74-86
2293	tsp_3	Thrombospondin type 3 repeat	0.0015	14.6	6	114-126
2293	tsp_3	Thrombospondin type 3 repeat	0.03	10.3	7	127-142
2293	TSPC	Thrombospondin C-terminal region	7.1e-176	594.4	1	167-367
2293	Mnd1	Mnd1 family	0.68	3.4	1	366-374
2294	Vps52	Vps52 / Sac2 family	0.087	3.9	1	154-183
2294	Complex1_17_2kD	NADH:ubiquinone oxidoreductase 17.2 k	0.25	6.1	1	562-587
2294	mRNA_triPase	mRNA capping enzyme, beta chain	0.33	4.0	1	934-966
2294	DUF424	Protein of unknown function (DUF424)	0.79	4.6	1	1002-1017
2295	sodcu	Copper/zinc superoxide dismutase (SOD	1	2.0	1	10-23
2295	DapB_C	Dihydrodipicolinate reductase, C-term	0.84	4.5	1	17-31
2295	PTR2	POT family	1.9e-104	357.1	1	82-475
2296	FH2	Formin Homology 2 Domain	0.0052	11.5	1	98-144

570

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2296	NTP_transf_2	Nucleotidyltransferase domain	0.049	7.4	1	104-170
2296	FH2	Formin Homology 2 Domain	1.6e-05	20.9	2	158-186
2297	zf-C2H2	Zinc finger, C2H2 type	0.01	15.5	1	1-21
2297	XPA_N	XPA protein N-terminal	0.51	5.7	1	24-36
2297	TFIIS	1/6 1 9 [. 31 39	0.16	7.4	2	27-37
2297	zf-C2H2	Zinc finger, C2H2 type	6.7e-06	28.3	2	27-49
2297	XPA_N	XPA protein N-terminal	0.49	5.8	2	52-64
2297	TFIIS	1/6 1 9 [. 31 39	0.18	7.2	3	55-65
2297	zf-C2H2	Zinc finger, C2H2 type	4e-06	29.2	3	55-77
2297	TFIIS	1/6 1 9 [. 31 39	0.51	5.7	4	83-93
2297	zf-C2H2	Zinc finger, C2H2 type	2.7e-05	25.9	4	83-105
2297	zf-C2H2	Zinc finger, C2H2 type	7.8e-07	32.1	5	111-133
2297	XPA_N	4/5 108 120 .. 1 13	0.45	5.9	5	136-148
2297	eIF5_eIF2B	Domain found in IF2B/IF5	0.95	3.5	1	139-149
2297	TFIIS	1/6 1 9 [. 31 39	0.069	8.7	6	139-149
2297	Transposase_12	Transposase	0.48	3.6	1	139-165
2297	zf-C2H2	Zinc finger, C2H2 type	6.6e-07	32.4	6	139-161
2298	Sprouty	Sprouty protein (Spry)	1.2e-17	55.0	1	70-107
2299	HAMP	HAMP domain	0.21	7.3	1	9-42
2299	PA	PA domain	3.6e-19	65.4	1	155-255
2299	Peptidase_M28	Peptidase family M28	2e-118	403.6	1	332-585
2299	Borrelia_lipo	Borrelia burgdorferi virulent strain	0.98	2.5	1	591-604
2299	TFR_dimer	Transferrin receptor-like dimerisatio	1e-65	228.5	1	597-739
2300	GvpG	Gas vesicle protein G	0.088	6.7	1	43-75
2301	Sema	Sema domain	5.5e-05	17.6	1	34-113
2303	ZZ	Zinc finger, ZZ type	1e-12	48.2	1	5-50
2303	SoxD	Sarcosine oxidase, delta subunit fami	0.97	4.2	1	79-86
2303	zf-C2H2	Zinc finger, C2H2 type	0.00067	20.3	1	80-103
2303	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.3	3.6	1	95-115
2303	SPDY	Domain of unknown function (DUF317)	0.6	4.4	1	119-133
2303	Di19	Drought induced 19 protein (Di19)	0.00056	13.0	1	314-330
2305	ig	Immunoglobulin domain	1.7e-05	23.4	1	37-114
2305	Phage_cap_E	Phage major capsid protein E	0.79	2.8	1	128-137
2306	MHC_I	Class I Histocompatibility antigen, d	9.2e-142	481.1	1	32-210
2306	DUF497	Protein of unknown function (DUF497)	0.2	6.7	1	50-63
2306	ig	Immunoglobulin domain	7.9e-09	35.9	1	227-292
2306	DUF395	YeeE/YedE family (DUF395)	0.19	7.2	1	317-342
2307	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	5.8e-05	18.1	1	15-98
2307	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	0.7	3.5	2	113-138
2308	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	5.8e-05	18.1	1	15-98
2308	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	0.7	3.5	2	113-138
2309	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	5.8e-05	18.1	1	15-98
2309	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	0.7	3.5	2	113-138
2310	LBP_BPI_CETP_C	LBP / BPI / CETP family, C-terminal do	5.8e-05	18.1	1	15-98
2310	LBP_BPI_CETP	LBP / BPI / CETP family, C-terminal	0.7	3.5	2	113-138

571

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
	C	do				
2311	Secretogranin_V	Neuroendocrine protein 7B2 precursor	2.8e-136	462.9	1	10-214
2312	Cyto_heme_lyase	Cytochrome c/c1 heme lyase	0.82	1.8	1	97-120
2313	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	6.9e-46	159.3	1	8-185
2313	Acyl_transf_3	Acyltransferase family	0.12	6.3	1	110-155
2314	Cna_B	Cna protein B-type domain	0.17	5.9	1	52-85
2314	PDZ	PDZ domain (Also known as DHR or GLGF	8.1e-12	43.4	1	52-130
2315	PID	Phosphotyrosine interaction domain (PT	3.3e-47	160.5	1	46-172
2317	pkinase	Protein kinase domain	8.3e-74	255.4	1	22-282
2318	lipocalin	Lipocalin / cytosolic fatty-acid binding p	2.3e-42	150.9	1	58-206
2318	Triabin	Triabin	0.0018	12.1	1	139-156
2319	lactamase_B	Metallo-beta-lactamase superfamily	2.3e-06	24.6	1	26-74
2320	annexin	Annexin	2.5e-05	21.2	1	1-20
2320	annexin	Annexin	1.1e-29	107.6	2	26-92
2320	annexin	Annexin	9.7e-28	100.7	3	109-176
2320	annexin	Annexin	2.8e-33	120.4	4	185-251
2321	SNF	Sodium:neurotransmitter symporter fam	9.5e-260	873.0	1	38-417
2321	ATP-sulfurylase	ATP-sulfurylase	0.28	3.8	1	42-64
2321	DUF900	Protein of unknown function (DUF900)	0.98	2.8	1	251-263
2323	Glypican	Glypican	2.4e-60	201.2	1	3-115
2324	PAP_assoc	PAP/25A associated domain	1.6e-14	51.8	1	274-333
2324	Isochorismatase	Isochorismatase family	0.49	4.1	1	484-520
2326	Sec23_trunk	Sec23/Sec24 trunk domain	0.47	4.0	1	22-33
2326	Hydrolase	haloacid dehalogenase-like hydrolase	0.77	3.7	1	26-56
2327	Sec23_trunk	Sec23/Sec24 trunk domain	0.47	4.0	1	22-33
2327	Hydrolase	haloacid dehalogenase-like hydrolase	0.77	3.7	1	26-56
2328	A2M	Alpha-2-macroglobulin family	6.3e-23	75.5	1	4-86
2329	A2M	Alpha-2-macroglobulin family	6.3e-23	75.5	1	4-86
2330	A2M	Alpha-2-macroglobulin family	6.3e-23	75.5	1	4-86
2331	A2M	Alpha-2-macroglobulin family	6.3e-23	75.5	1	4-86
2332	A2M	Alpha-2-macroglobulin family	6.3e-23	75.5	1	4-86
2333	COesterase	Carboxylesterase	4.3e-42	142.8	1	8-142
2333	A2M_N	Alpha-2-macroglobulin family N-termina	0.83	2.3	1	12-28
2334	EGF	EGF-like domain	0.017	11.7	1	5-26
2334	TIL	Trypsin Inhibitor like cysteine rich doma	0.85	3.5	1	10-26
2336	Corona_NS4	Coronavirus non-structural protein NS	0.47	3.5	1	20-43
2336	ig	Immunoglobulin domain	0.052	10.4	1	57-119
2336	fn3	Fibronectin type III domain	2.4e-16	58.5	1	145-231
2339	Keratin_B2	Keratin, high sulfur B2 protein	1e-19	69.2	1	21-145
2340	Keratin_B2	Keratin, high sulfur B2 protein	1e-19	69.2	1	21-145
2341	Keratin_B2	Keratin, high sulfur B2 protein	1e-19	69.2	1	21-145
2342	abhydro_lipase	ab-hydrolase associated lipase region	1.9e-32	117.8	1	87-157
2342	abhydrolase	alpha/beta hydrolase fold	9.5e-19	67.6	1	171-448
2343	abhydro_lipase	ab-hydrolase associated lipase region	1.9e-32	117.8	1	87-157
2343	abhydrolase	alpha/beta hydrolase fold	9.5e-19	67.6	1	171-448
2344	7tm_3	7 transmembrane receptor (metabotropic	0.75	3.2	1	26-47

572
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2344	7tm_3	7 transmembrane receptor (metabotropic)	0.00057	14.4	2	67-111
2344	Condensation	Condensation domain	0.36	4.2	1	159-171
2344	7tm_3	7 transmembrane receptor (metabotropic)	1.4e-05	20.1	3	170-273
2345	GASA	Gibberellin regulated protein	0.35	1.3	1	25-54
2345	lectin_c	Lectin C-type domain	1.3e-25	95.2	1	103-216
2346	GASA	Gibberellin regulated protein	0.35	1.3	1	25-54
2346	lectin_c	Lectin C-type domain	1.3e-25	95.2	1	103-216
2347	7tm_1	7 transmembrane receptor (rhodopsin f	2.6e-50	149.5	1	99-348
2347	endotoxin_N	delta endotoxin, N-terminal domain	0.87	3.6	1	253-283
2347	Pox_D2	Pox virus D2 protein	0.93	1.2	1	366-379
2347	7tm_1	7 transmembrane receptor (rhodopsin f	9.6e-48	141.7	2	417-666
2347	PAZ	PAZ domain	0.48	4.7	1	540-567
2348	Hanta_G2	Hantavirus glycoprotein G2	0.098	4.8	1	84-112
2350	An_peroxidase	Animal haem peroxidase	1.3e-91	311.6	1	2-232
2350	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	24-32
2350	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	117-134
2351	An_peroxidase	Animal haem peroxidase	1.3e-91	311.6	1	2-232
2351	7tm_1	7 transmembrane receptor (rhodopsin f	0.22	2.7	1	24-32
2351	Peptidase_C1	Papain family cysteine protease	0.76	2.1	1	117-134
2352	Arch fla_DE	Archaeal flagella protein	0.42	5.0	1	86-99
2353	UBX	UBX domain	0.36	6.0	1	141-159
2353	FTCD_C	Formiminotransferase-cyclodeaminase	0.21	6.0	1	188-218
2353	3H	3H domain	0.46	6.2	1	248-260
2354	Torsin	Torsin	3e-189	638.8	1	17-288
2354	2_5_ligase	2',5' RNA ligase family	0.13	7.6	1	101-133
2354	DUF254	SAND family protein	0.22	3.0	1	110-129
2355	SPX	SPX domain	0.84	1.0	1	66-86
2359	DUF895	Eukaryotic protein of unknown function	0.68	4.1	1	183-199
2360	RWD	RWD domain	9.8e-40	142.2	1	17-131
2360	globin	Globin	0.048	8.6	1	94-126
2360	eRF1_2	eRF1 domain 2	0.72	4.1	1	120-133
2360	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	7.5e-10	27.9	1	141-207
2360	DNA_ligase_ZB D	NAD-dependent DNA ligase C4 zinc finge	0.37	5.8	1	202-213
2360	zf-MIZ	MIZ zinc finger	0.28	4.6	1	203-213
2360	ApoA-II	Apolipoprotein A-II (ApoA-II)	0.94	3.6	1	267-278
2361	TMS_TDE	TMS membrane protein/tumour differentia	0.048	5.2	1	33-63
2361	sugar_tr	Sugar (and other) transporter	1.5e-05	19.2	1	65-138
2362	ig	Immunoglobulin domain	0.00079	17.2	1	42-98
2363	ig	Immunoglobulin domain	0.00079	17.2	1	42-98
2364	IBN_NT	Importin-beta N-terminal domain	6.3e-16	58.6	1	65-109
2365	HIM	Haemagglutinin motif	0.18	7.7	1	375-391
2365	Fascin	Fascin protein	0.29	0.7	1	808-818
2370	DUF350	Domain of Unknown Function (DUF350)	0.86	5.1	1	106-128
2372	Apolipoprotein	Apolipoprotein A1/A4/E family	0.95	3.4	1	54-78
2372	F5_F8_type_C	F5/8 type C domain	2.2e-36	111.6	1	84-171
2373	Apolipoprotein	Apolipoprotein A1/A4/E family	0.95	3.4	1	54-78
2373	F5_F8_type_C	F5/8 type C domain	2.2e-36	111.6	1	84-171
2374	Ribosomal_S3_C	Ribosomal protein S3, C-terminal	0.98	3.4	1	65-71

573

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
		domai				
2374	Ricin_B_lectin	QXW lectin repeat	0.13	8.4	1	175-204
2374	Ricin_B_lectin	QXW lectin repeat	0.00073	16.5	2	266-304
2375	pkinase	Protein kinase domain	1.2e-11	41.7	1	38-119
2377	PH	PH domain	0.023	9.0	1	43-81
2377	BTK	BTK motif	1.9e-06	26.9	1	105-141
2379	EGF	EGF-like domain	0.047	10.1	1	1-17
2379	EGF	EGF-like domain	3.7e-07	28.6	2	33-68
2379	EB	EB module	0.73	4.1	1	39-68
2380	M_tail	Myosin tail	0.32	4.7	1	192-222
2380	SurE	Survival protein SurE	0.68	2.6	1	317-330
2380	Pox_A11	Poxvirus A11 Protein	0.17	3.2	1	364-382
2385	7tm_1	7 transmembrane receptor (rhodopsin f	8.6e-47	138.9	1	83-332
2386	7tm_1	7 transmembrane receptor (rhodopsin f	8.6e-47	138.9	1	83-332
2388	Calcyon	D1 dopamine receptor-interacting protein	3.7e-42	136.9	1	1-66
2389	IL1	Interleukin-1 / 18	2.6e-23	83.4	1	79-180
2390	filament	Intermediate filament protein	7.3e-68	235.6	1	2-189
2390	K-box	K-box region	0.11	7.0	1	11-29
2390	bZIP	bZIP transcription factor	0.2	7.0	1	49-86
2390	Ribosomal_L29	Ribosomal L29 protein	0.71	5.2	1	104-130
2392	MgpC	MgpC protein precursor	0.99	2.8	1	1-27
2393	filament	Intermediate filament protein	1.8e-14	54.6	1	9-80
2394	Peptidase_M10	Matrixin	4.5e-47	166.6	1	11-103
2394	Peptidase_M10	Matrixin	9.2e-17	64.0	2	107-145
2395	Peptidase_M10	Matrixin	4.5e-47	166.6	1	11-103
2395	Peptidase_M10	Matrixin	9.2e-17	64.0	2	107-145
2396	SUFU	Suppressor of fused protein (SUFU)	0	1218.3	1	3-484
2397	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal doma	3.4e-20	69.4	1	38-103
2398	MCPsignal	Methyl-accepting chemotaxis protein (0.21	4.1	1	363-379
2399	MCPsignal	Methyl-accepting chemotaxis protein (0.21	4.1	1	363-379
2402	DUF846	Eukaryotic protein of unknown functio	4.3e-05	15.0	1	63-93
2405	UK	Virulence determinant	0.083	7.0	1	48-72
2405	TIP120	TBP (TATA-binding protein) -interacti	0	3271.6	1	59-1252
2405	HEAT	HEAT repeat	0.093	8.3	2	282-320
2405	HEAT	HEAT repeat	0.04	9.5	3	377-398
2405	Armadillo_seg	Armadillo/beta-catenin-like repeat	0.2	8.0	2	716-755
2406	lectin_c	Lectin C-type domain	2.5e-16	64.4	1	168-274
2412	PTE	Phosphotriesterase family	8.2e-207	697.2	1	38-380
2412	gntR	Bacterial regulatory proteins, gntR f	0.17	7.0	1	141-160
2414	filament	Intermediate filament protein	0.28	5.0	1	199-228
2414	Transposase_8	Transposase	0.57	5.0	1	200-220
2414	DUF972	Protein of unknown function (DUF972)	0.76	4.2	1	201-245
2414	Rop	Rop protein	0.55	3.6	1	242-249
2414	MoaE	MoaE protein	0.18	7.2	1	467-480
2414	WH2	WH2 motif	0.14	8.9	1	468-485
2419	Pox_int_trans	Poxvirus intermediate transcription fa	0.092	5.7	1	119-147
2419	ABA_WDS	ABA/WDS induced protein	0.81	4.5	1	185-201
2419	DUF738	Protein of unknown function (DUF738)	0.89	3.3	1	297-316
2419	IpaB_EvcA	IpaB/EvcA family	0.65	3.8	1	460-485

574

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2420	PS_Dcarboxylase	Phosphatidylserine decarboxylase	3.9e-60	209.9	1	80-323
2422	DUF199	Uncharacterized BCR, COG1481	0.5	3.4	1	138-154
2423	ADP_PFK_GK	ADP-specific Phosphofructokinase/Gluco	4.1e-208	701.5	1	6-408
2423	Mannitol_dh	Mannitol dehydrogenase	0.052	6.8	1	310-329
2424	ASFV_L11L	African swine fever virus (ASFV) L11L	0.89	3.1	1	8-19
2428	Ribosomal_S25	S25 ribosomal protein	0.00053	15.1	1	12-43
2429	ank	Ankyrin repeat	0.011	12.7	1	1-18
2429	ank	Ankyrin repeat	0.0036	14.4	2	19-51
2429	ank	Ankyrin repeat	1.5e-11	44.5	3	52-84
2429	ank	Ankyrin repeat	1.2e-08	34.1	4	85-117
2429	ank	Ankyrin repeat	3.3e-08	32.5	5	118-150
2429	ank	Ankyrin repeat	3.4e-11	43.2	6	151-183
2429	ank	Ankyrin repeat	1.3e-08	33.9	7	184-217
2429	ank	Ankyrin repeat	0.0027	14.9	8	218-250
2429	ank	Ankyrin repeat	8.5e-08	31.1	9	251-283
2429	ank	Ankyrin repeat	0.013	12.4	10	284-308
2429	ank	Ankyrin repeat	8.3e-08	31.1	11	335-367
2429	ank	Ankyrin repeat	1.1e-09	37.8	12	368-400
2429	ank	Ankyrin repeat	6.9e-07	27.8	13	401-461
2429	endonuclease_7	Recombination endonuclease VII	0.034	9.6	1	417-441
2429	ank	Ankyrin repeat	0.0047	14.0	14	462-485
2430	trypsin	Trypsin	1.5e-23	72.8	1	61-237
2430	PDZ	PDZ domain (Also known as DHR or GLGF)	5.1e-08	30.0	1	285-339
2431	vwc	von Willebrand factor type C domain	5.4e-05	19.0	1	66-105
2432	MethyltransfD12	D12 class N6 adenine-specific DNA met	1.4e-36	128.4	1	39-163
2432	Ribosomal_L1	Ribosomal protein L1p/L10e family	0.57	1.6	1	99-117
2433	lipocalin	Lipocalin / cytosolic fatty-acid binding pr	4.7e-11	42.2	1	12-80
2434	tRNA_anti	OB-fold nucleic acid binding domain	2.7e-15	56.4	1	71-145
2434	tRNA-synt_2	tRNA synthetases class II (D, K and N	2.2e-59	207.4	1	162-410
2434	Transglutamin_C	Transglutaminase family, C-terminal i	0.79	4.1	1	229-256
2434	RNA_helicase	RNA helicase	0.13	5.7	1	266-308
2435	FAD_binding_2	FAD binding domain	1.6e-53	181.4	1	22-117
2436	RasGEF	RasGEF domain	6.8e-18	69.6	1	35-115
2437	KH	KH domain	3.8e-17	61.6	1	78-124
2437	Peripla_BP_2	Periplasmic binding protein	0.71	3.5	1	116-132
2437	KH	KH domain	2.4e-10	38.1	2	162-189
2439	transket_pyr	Transketolase, pyridine binding domai	1.6e-51	176.9	1	76-254
2439	DUF924	Bacterial protein of unknown function	0.88	2.8	1	77-98
2439	Indigoidine_A	Indigoidine synthase A like protein	0.51	4.4	1	233-247
2439	transketolase_C	Transketolase, C-terminal domain	9.7e-42	137.2	1	272-398
2440	Calsequestrin	Calsequestrin	8.5e-292	979.5	1	42-427
2440	thiore	Thioredoxin	0.057	9.0	1	160-189
2441	Bacillus_PapR	Bacillus PapR protein	0.68	3.6	1	62-77
2441	arf	ADP-ribosylation factor family	0.76	2.9	1	290-312
2441	RNA_capsid	Calicivirus putative RNA polymerase/ca	0.65	1.9	1	346-355
2442	ig	Immunoglobulin domain	2.7e-09	37.6	1	35-112
2442	Relaxase	Relaxase/Mobilisation nuclease domain	0.91	3.4	1	71-90

575
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2442	virus_P-coat	Viral coat protein	0.7	4.0	1	140-152
2442	ig	Immunoglobulin domain	9.8e-07	28.1	2	156-230
2445	cadherin	Cadherin domain	0.045	8.9	1	77-106
2445	cadherin	Cadherin domain	1.4e-14	52.6	2	150-245
2445	cadherin	Cadherin domain	4.8e-24	85.5	3	259-350
2445	cadherin	Cadherin domain	3.6e-14	51.1	4	364-455
2445	cadherin	Cadherin domain	2.9e-22	79.4	5	469-565
2445	Hema_HEFG	Hemagglutinin domain of haemagglutini	0.25	4.6	1	515-531
2445	cadherin	Cadherin domain	8.2e-14	49.9	6	594-677
2447	ShTK	ShTK domain	0.92	3.9	1	74-82
2448	zf-C2H2	Zinc finger, C2H2 type	0.00076	20.0	2	50-73
2448	zf-C2H2	Zinc finger, C2H2 type	0.036	13.3	3	80-106
2448	zf-C2H2	Zinc finger, C2H2 type	0.00095	19.6	5	198-221
2450	Alpha_L_fucos	Alpha-L-fucosidase	0.018	8.4	1	10-34
2451	TCTP	Translationally controlled tumour protein	7.3e-13	42.0	1	20-54
2452	Herpes_gG	Glycoprotein GG/GX	0.39	2.9	1	29-54
2452	Osteopontin	Osteopontin	1.1e-128	410.7	1	42-218
2452	Flu_M1	Influenza Matrix protein (M1)	1	3.0	1	52-65
2453	serpin	Serpin (serine protease inhibitor)	0.99	2.0	1	68-92
2454	HATPase_c	Histidine kinase-, DNA gyrase B-, and	3.8e-15	54.5	1	92-240
2454	Pox_N2L	Poxvirus N2L protein	0.18	5.3	1	162-176
2454	DNA_gyraseB	DNA gyrase B	4.1e-57	199.9	1	286-446
2454	FokI_N	Restriction endonuclease FokI, recogn	0.12	6.3	1	530-539
2454	DNA_topoisoIV	DNA gyrase/topoisomerase IV, subunit	7.6e-190	610.8	1	729-1196
2454	DUF188	Uncharacterized BCR, YaiI/Yqx D family	0.025	8.2	1	1171-1197
2456	PCI	PCI domain	0.27	6.1	1	66-93
2457	C1q	C1q domain	6.3e-06	23.8	1	98-138
2458	BTB	BTB/POZ domain	1.9e-17	64.4	1	62-124
2459	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal doma	3.4e-20	69.4	1	38-103
2460	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal doma	3.4e-20	69.4	1	38-103
2461	LBP_BPI_CETP	LBP / BPI / CETP family, N-terminal doma	3.4e-20	69.4	1	38-103
2462	DUF408	Domain of Unknown Function (DUF408)	1e-11	42.8	1	1-43
2462	DUF584	Protein of unknown function, DUF584	0.67	2.3	1	224-250
2464	Securin	Securin sister-chromatid separation inhibito	1	2.9	1	19-34
2465	Nuf2	Nuf2 family	3.3e-104	356.3	1	6-153
2465	Corona_NS2A	Coronavirus NS2A protein	0.42	2.3	1	133-139
2465	Syntaxin	Syntaxin	0.31	6.3	1	142-242
2465	HR1	Hr1 repeat	0.099	7.1	1	192-219
2465	LEA	Late embryogenesis abundant protein	0.79	5.0	1	254-279
2465	Mob_Pre	Plasmid recombination enzyme	0.97	1.9	1	366-376
2465	G-gamma	GGL domain	0.08	6.9	1	403-424
2465	OKR_DC_1_N	Orn/Lys/Arg decarboxylase, N-terminal	0.19	4.3	1	426-450

576

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2466	pkinase	Protein kinase domain	0.036	7.9	1	62-123
2466	HEAT	HEAT repeat	0.13	7.8	1	335-373
2467	UreE_C	UreE urease accessory protein, C-termi	0.09	7.9	1	119-140
2467	Pox_A_type_inc	Viral A-type inclusion protein repeat	0.59	6.3	1	194-215
2470	DUF563	Protein of unknown function (DUF563)	0.86	2.4	1	36-47
2471	Fumarate_red_D	Fumarate reductase subunit D	0.28	5.3	1	79-103
2473	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	2.9e-06	17.8	1	59-96
2473	IBR	IBR domain	0.0064	10.5	1	70-114
2473	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	0.029	6.4	2	124-144
2474	Aa_trans	Transmembrane amino acid transporter p	1.4e-41	148.3	1	72-276
2475	Aa_trans	Transmembrane amino acid transporter p	1.4e-41	148.3	1	72-276
2477	LRR	Leucine Rich Repeat	0.093	8.7	1	30-51
2477	LRR	Leucine Rich Repeat	0.57	6.0	2	56-77
2478	SAB	SAB domain	0.65	5.3	1	57-82
2478	RPE65	Retinal pigment epithelial membrane prote	4.3e-27	91.2	1	85-202
2479	SAB	SAB domain	0.65	5.3	1	57-82
2479	RPE65	Retinal pigment epithelial membrane prote	4.3e-27	91.2	1	85-202
2480	Pox_A_type_inc	Viral A-type inclusion protein repeat	0.42	6.8	1	18-40
2480	spectrin	Spectrin repeat	0.61	4.8	1	18-45
2481	CD34_antigen	CD34 antigen protein	0.88	0.3	1	6-34
2481	DUF999	Protein of unknown function (DUF999)	0.28	5.1	1	14-35
2481	BCL_N	BCL7, N-terminal conserver region	0.39	6.1	1	174-195
2481	serpin	Serpin (serine protease inhibitor)	3.7e-24	81.8	1	431-546
2482	PMP22_Claudin	PMP-22/EMP/MP20/Claudin family	3.7e-89	306.4	1	40-218
2482	mce	mce related protein	0.74	4.3	1	159-179
2483	PAP2	PAP2 superfamily	4.5e-15	54.4	1	24-159
2488	SMC_C	SMC family, C-terminal domain	1.6e-17	60.6	1	418-475
2488	SMC_C	SMC family, C-terminal domain	1.1e-43	150.2	2	477-540
2488	Armadillo_seg	Armadillo/beta-catenin-like repeat	4.6e-14	53.0	2	551-591
2488	Armadillo_seg	Armadillo/beta-catenin-like repeat	1.4e-08	33.5	3	594-634
2489	IER	Immediate early response protein (IER)	0.063	3.8	1	194-206
2490	disintegrin	Disintegrin	3.3e-36	123.1	1	4-79
2490	EGF	EGF-like domain	0.0023	14.8	1	231-258
2499	ARPF	Aromatic-Rich Protein Family	1.4e-10	36.3	1	89-234
2502	Paf1	Paf1	2.3e-17	65.0	1	1-61
2502	ig	Immunoglobulin domain	0.015	12.4	1	68-113
2502	ank	Ankyrin repeat	0.025	11.4	1	186-204
2504	PH	PH domain	0.028	8.7	1	61-153
2504	DAGKc	Diacylglycerol kinase catalytic domain	0.00051	15.4	1	161-213
2505	ig	Immunoglobulin domain	0.069	9.9	1	48-120
2505	ig	Immunoglobulin domain	6.5e-09	36.2	2	161-219
2506	ig	Immunoglobulin domain	0.069	9.9	1	48-120
2506	ig	Immunoglobulin domain	6.5e-09	36.2	2	161-219
2508	7tm_1	7 transmembrane receptor (rhodopsin famil	8.2e-29	85.0	1	49-179
2508	7tm_1	7 transmembrane receptor (rhodopsin famil	1.6e-13	39.1	2	210-267
2517	Acyl-CoA_dh_M	Acyl-CoA dehydrogenase, middle domain	0.0071	11.7	1	99-136
2517	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal	6.7e-50	175.9	1	415-566

577
TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
		doma				
2518	Acyl-CoA_dh_M	Acyl-CoA dehydrogenase, middle domain	0.0071	11.7	1	99-136
2518	Acyl-CoA_dh	Acyl-CoA dehydrogenase, C-terminal doma	6.7e-50	175.9	1	415-566
2519	Cation_efflux	Cation efflux family	3e-09	34.4	1	33-109
2520	CaMBD	Calmodulin binding domain	0.074	7.8	1	451-467
2520	IQ	IQ calmodulin-binding motif	1.3e-05	22.1	2	473-493
2520	IQ	IQ calmodulin-binding motif	1.6e-05	21.8	3	532-552
2524	PAP2	PAP2 superfamily	7.6e-19	67.8	1	39-151
2525	PAP2	PAP2 superfamily	7.6e-19	67.8	1	39-151
2529	LRR	Leucine Rich Repeat	0.00025	17.3	1	201-226
2529	LRR	Leucine Rich Repeat	0.0019	14.3	2	227-246
2529	LRR	Leucine Rich Repeat	0.13	8.1	4	272-297
2529	LRR	Leucine Rich Repeat	0.00025	17.3	5	298-317
2529	LRR	Leucine Rich Repeat	5.2e-05	19.6	6	319-342
2529	LRR	Leucine Rich Repeat	0.37	6.6	7	343-368
2530	ig	Immunoglobulin domain	0.26	7.8	1	55-122
2530	ig	Immunoglobulin domain	0.0043	14.5	2	162-220
2530	ig	Immunoglobulin domain	0.00023	19.2	3	267-321
2531	ig	Immunoglobulin domain	0.26	7.8	1	55-122
2531	ig	Immunoglobulin domain	0.0043	14.5	2	162-220
2531	ig	Immunoglobulin domain	0.00023	19.2	3	267-321
2532	tsp_1	Thrombospondin type 1 domain	2.9e-07	25.9	1	59-103
2533	Guanylin	Guanylin precursor	0.0007	9.1	1	12-35
2533	Apo-CII	Apolipoprotein C-II	3.4e-57	200.2	1	34-111
2534	Guanylin	Guanylin precursor	0.0007	9.1	1	12-35
2534	Apo-CII	Apolipoprotein C-II	3.4e-57	200.2	1	34-111
2536	zf-C2H2	Zinc finger, C2H2 type	0.0012	19.3	1	279-301
2536	zf-C2H2	Zinc finger, C2H2 type	2.2e-06	30.3	2	307-329
2536	zf-C2H2	Zinc finger, C2H2 type	0.086	11.7	3	337-355
2540	FA_desaturase	Fatty acid desaturase	5.1e-42	145.2	1	8-159
2541	rnaseH	RNase H	7.1e-16	53.4	1	86-184
2541	MutS_III	MutS domain III	4.2e-06	22.9	1	253-277
2541	MutS_V	MutS domain V	6e-164	543.6	1	282-517
2542	ig	Immunoglobulin domain	7.6e-08	32.2	1	1-57
2542	fn3	Fibronectin type III domain	2.8e-16	58.3	1	79-165
2543	MAM	MAM domain	1.5e-43	154.8	1	3-102
2544	kazal	Kazal-type serine protease inhibitor domain	7.7e-06	25.8	1	40-87
2544	ig	Immunoglobulin domain	4.1e-07	29.5	1	105-174
2545	RNA_helicase	RNA helicase	0.031	7.9	1	85-112
2545	ATP-bind	Conserved hypothetical ATP binding prote	0.055	7.3	1	90-103
2547	ig	Immunoglobulin domain	0.015	12.4	1	10-28
2547	ig	Immunoglobulin domain	0.098	9.4	2	72-98
2549	serpin	Serpin (serine protease inhibitor)	5.4e-18	60.8	1	68-112
2551	DREV	DREV methyltransferase	7.3e-233	680.7	1	57-318
2553	ank	Ankyrin repeat	1.8e-07	29.9	2	44-76
2553	ank	Ankyrin repeat	0.026	11.4	3	77-102
2554	pkinase	Protein kinase domain	3.5e-64	223.4	1	117-375
2555	pkinase	Protein kinase domain	3.5e-64	223.4	1	117-375
2557	tRNA-synt_1e	tRNA synthetases class I (C)	0.0002	14.0	1	99-129

578

TABLE 4B

SEQ ID	Model	Description	E_value	Score	Repeats	Position
2557	tRNA-synt_1	tRNA synthetases class I (I, L, M and V)	2.4e-07	23.7	1	99-137
2558	MHC_II_beta	Class II histocompatibility antigen, beta	1.4e-43	149.3	1	41-116
2562	fn3	Fibronectin type III domain	0.0065	11.9	1	18-105
2563	A2M	Alpha-2-macroglobulin family	6.3e-23	75.5	1	4-86

TABLE 5

SEQ ID	Position	Maximum score	Average score
685	1-26	0.982	0.908
689	1-19	0.975	0.888
691	1-49	0.944	0.603
695	1-24	0.993	0.943
697	1-26	0.919	0.670
698	1-20	0.988	0.939
706	1-20	0.989	0.973
707	1-24	0.973	0.922
710	1-33	0.957	0.789
712	1-57	0.975	0.488
714	1-42	0.958	0.680
715	1-42	0.958	0.687
725	1-18	0.978	0.956
728	1-22	0.980	0.917
732	1-27	0.974	0.932
733	1-27	0.974	0.932
734	1-27	0.974	0.932
738	1-75	0.923	0.462
742	1-23	0.905	0.707
744	1-33	0.981	0.884
747	1-20	0.991	0.954
748	1-30	0.950	0.785
753	1-30	0.991	0.936
754	1-17	0.978	0.905
755	1-16	0.967	0.933
756	1-18	0.970	0.897
757	1-17	0.948	0.869
758	1-17	0.948	0.869
759	1-21	0.916	0.820
762	1-14	0.972	0.951
781	1-38	0.917	0.618
784	1-21	0.984	0.869
796	1-19	0.982	0.959
797	1-19	0.982	0.959
798	1-19	0.982	0.959
800	1-65	0.857	0.487
801	1-45	0.903	0.565
803	1-36	0.985	0.834
804	1-21	0.993	0.855
806	1-20	0.937	0.779
807	1-20	0.937	0.779
808	1-20	0.937	0.779
809	1-32	0.972	0.885
811	1-25	0.991	0.948
814	1-28	0.948	0.827
815	1-33	0.947	0.744
816	1-23	0.986	0.908
817	1-23	0.986	0.908
819	1-21	0.959	0.755
825	1-35	0.974	0.637

TABLE 5

SEQ ID	Position	Maximum score	Average score
826	1-42	0.981	0.909
834	1-21	0.978	0.751
835	1-44	0.985	0.831
836	1-44	0.985	0.814
838	1-31	0.986	0.935
844	1-18	0.951	0.879
845	1-18	0.951	0.879
848	1-20	0.992	0.794
852	1-24	0.976	0.901
853	1-24	0.976	0.901
855	1-25	0.933	0.751
858	1-24	0.915	0.567
867	1-17	0.968	0.863
868	1-17	0.968	0.863
869	1-34	0.987	0.781
870	1-16	0.901	0.686
872	1-14	0.964	0.931
877	1-21	0.988	0.958
878	1-22	0.915	0.833
879	1-25	0.922	0.765
880	1-25	0.922	0.765
882	1-20	0.917	0.819
888	1-24	0.985	0.945
889	1-17	0.989	0.945
890	1-23	0.995	0.938
891	1-24	0.971	0.882
893	1-16	0.891	0.770
894	1-20	0.972	0.859
900	1-22	0.931	0.862
901	1-24	0.993	0.937
907	1-22	0.974	0.850
908	1-23	0.993	0.950
909	1-15	0.994	0.617
910	1-23	0.993	0.950
919	1-15	0.947	0.797
924	1-19	0.964	0.927
925	1-19	0.964	0.927
927	1-26	0.962	0.783
930	1-43	0.987	0.765
932	1-31	0.992	0.803
934	1-23	0.984	0.884
936	1-48	0.967	0.624
939	1-30	0.973	0.851
941	1-18	0.978	0.957
942	1-21	0.978	0.937
948	1-21	0.965	0.760
951	1-29	0.989	0.946
954	1-31	0.945	0.587
956	1-22	0.836	0.491
958	1-28	0.984	0.903

TABLE 5

SEQ ID	Position	Maximum score	Average score
960	1-24	0.987	0.924
961	1-24	0.987	0.924
962	1-24	0.987	0.924
965	1-21	0.993	0.934
966	1-43	0.974	0.653
967	1-32	0.953	0.778
968	1-40	0.972	0.632
970	1-24	0.981	0.938
971	1-24	0.981	0.776
973	1-28	0.923	0.694
978	1-37	0.968	0.746
979	1-37	0.968	0.746
980	1-23	0.984	0.943
981	1-18	0.961	0.869
982	1-24	0.971	0.865
983	1-21	0.988	0.937
984	1-20	0.938	0.716
985	1-20	0.938	0.716
986	1-25	0.913	0.560
988	1-16	0.969	0.949
993	1-39	0.972	0.817
994	1-21	0.970	0.808
996	1-22	0.977	0.837
1006	1-35	0.967	0.668
1010	1-24	0.980	0.902
1013	1-24	0.987	0.903
1014	1-24	0.987	0.903
1017	1-23	0.932	0.654
1019	1-20	0.984	0.868
1023	1-25	0.948	0.735
1024	1-23	0.968	0.924
1027	1-25	0.956	0.848
1028	1-16	0.993	0.980
1029	1-16	0.993	0.980
1031	1-33	0.985	0.813
1039	1-46	0.982	0.666
1041	1-41	0.988	0.886
1046	1-24	0.991	0.940
1048	1-19	0.991	0.934
1052	1-21	0.991	0.903
1053	1-25	0.971	0.897
1054	1-24	0.975	0.932
1055	1-18	0.986	0.965
1057	1-18	0.978	0.887
1058	1-18	0.978	0.887
1060	1-26	0.987	0.917
1062	1-34	0.991	0.901
1066	1-31	0.992	0.741
1068	1-22	0.962	0.919
1072	1-22	0.986	0.943

TABLE 5

SEQ ID	Position	Maximum score	Average score
1073	1-23	0.974	0.799
1075	1-33	0.986	0.886
1076	1-23	0.969	0.696
1077	1-23	0.969	0.696
1078	1-17	0.978	0.905
1079	1-30	0.935	0.717
1080	1-17	0.978	0.905
1081	1-17	0.978	0.905
1082	1-17	0.978	0.905
1083	1-26	0.936	0.809
1084	1-23	0.993	0.907
1085	1-18	0.969	0.643
1092	1-19	0.937	0.713
1096	1-39	0.995	0.594
1097	1-39	0.995	0.594
1100	1-20	0.964	0.902
1101	1-23	0.993	0.950
1102	1-23	0.993	0.950
1105	1-21	0.987	0.963
1106	1-19	0.947	0.709
1111	1-13	0.911	0.718
1117	1-20	0.930	0.706
1118	1-16	0.964	0.790
1121	1-24	0.968	0.825
1123	1-20	0.991	0.881
1128	1-22	0.969	0.871
1129	1-25	0.985	0.864
1130	1-25	0.985	0.864
1131	1-20	0.958	0.893
1132	1-21	0.942	0.717
1134	1-24	0.976	0.925
1136	1-14	0.972	0.951
1137	1-19	0.960	0.901
1139	1-33	0.995	0.835
1140	1-30	0.993	0.853
1141	1-30	0.993	0.853
1143	1-35	0.974	0.637
1144	1-42	0.981	0.909
1145	1-21	0.975	0.874
1150	1-21	0.914	0.729
1152	1-17	0.990	0.973
1153	1-17	0.990	0.973
1155	1-23	0.965	0.907
1161	1-39	0.954	0.705
1162	1-45	0.929	0.575
1165	1-19	0.939	0.857
1167	1-25	0.951	0.619
1170	1-37	0.978	0.830
1172	1-16	0.957	0.870
1173	1-16	0.957	0.870

TABLE 5

SEQ ID	Position	Maximum score	Average score
1174	1-21	0.914	0.729
1178	1-25	0.980	0.925
1179	1-17	0.915	0.659
1181	1-22	0.950	0.719
1186	1-18	0.985	0.928
1192	1-18	0.960	0.803
1196	1-48	0.905	0.599
1199	1-20	0.988	0.955
1200	1-16	0.907	0.635
1205	1-25	0.974	0.781
1207	1-28	0.965	0.842
1208	1-23	0.965	0.693
1210	1-21	0.988	0.911
1213	1-31	0.940	0.696
1214	1-17	0.983	0.956
1218	1-23	0.996	0.969
1219	1-15	0.967	0.909
1221	1-16	0.978	0.938
1222	1-32	0.939	0.646
1223	1-23	0.982	0.945
1226	1-31	0.991	0.925
1228	1-32	0.953	0.778
1231	1-23	0.965	0.907
1232	1-23	0.965	0.907
1233	1-23	0.965	0.907
1235	1-21	0.873	0.596
1240	1-20	0.987	0.949
1241	1-22	0.994	0.890
1244	1-27	0.998	0.952
1245	1-27	0.998	0.952
1247	1-23	0.980	0.931
1253	1-17	0.945	0.731
1258	1-20	0.984	0.923
1259	1-32	0.956	0.757
1261	1-20	0.967	0.781
1262	1-18	0.961	0.886
1265	1-23	0.991	0.915
1266	1-23	0.991	0.915
1267	1-19	0.973	0.788
1268	1-34	0.988	0.888
1269	1-21	0.922	0.610
1271	1-23	0.910	0.653
1272	1-18	0.997	0.757
1275	1-29	0.989	0.943
1278	1-34	0.994	0.867
1279	1-15	0.983	0.957
1280	1-15	0.969	0.641
1281	1-36	0.916	0.620
1282	1-36	0.916	0.620
1283	1-36	0.896	0.584

TABLE 5

SEQ ID	Position	Maximum score	Average score
1287	1-18	0.836	0.471
1288	1-31	0.952	0.767
1290	1-22	0.962	0.904
1292	1-33	0.904	0.641
1293	1-33	0.904	0.641
1295	1-27	0.962	0.882
1297	1-30	0.995	0.964
1298	1-30	0.995	0.964
1300	1-25	0.998	0.961
1302	1-16	0.921	0.729
1303	1-24	0.991	0.913
1310	1-52	0.987	0.492
1311	1-19	0.903	0.592
1314	1-16	0.887	0.735
1315	1-27	0.911	0.682
1316	1-27	0.911	0.682
1317	1-25	0.987	0.924
1319	1-20	0.973	0.759
1320	1-20	0.968	0.733
1322	1-16	0.969	0.894
1323	1-16	0.969	0.894
1324	1-28	0.957	0.874
1325	1-17	0.972	0.946
1326	1-17	0.972	0.946
1327	1-18	0.905	0.593
1328	1-16	0.895	0.561
1329	1-17	0.978	0.896
1330	1-20	0.988	0.963
1333	1-24	0.985	0.965
1335	1-22	0.966	0.767
1343	1-32	0.954	0.675
1344	1-18	0.951	0.879
1345	1-30	0.978	0.901
1347	1-20	0.961	0.880
1348	1-18	0.978	0.940
1350	1-23	0.989	0.868
1352	1-23	0.993	0.883
1354	1-25	0.924	0.567
1358	1-18	0.993	0.909
1359	1-15	0.855	0.706
1360	1-31	0.985	0.908
1361	1-17	0.995	0.950
1362	1-17	0.995	0.950
1364	1-29	0.962	0.860
1366	1-17	0.978	0.905
1368	1-26	0.958	0.843

585

TABLE 6

SEQ ID	GENOMIC LOCATION
1	17
2	17
3	17
4	17
5	17
6	2
7	7
8	16
9	15
10	20p11.2-11.22.
11	4
12	12
13	13
14	14q24
15	18
16	4
17	12q
18	7
19	22q12.2
20	17
21	1
22	20
23	1
24	6
25	15
26	13q34
27	2q21.2
28	2q21.2
29	12
30	4
31	4
32	19
33	19-specific
34	3
35	2
36	1
37	11
38	17
39	10
41	Xp22
42	8
43	19
44	6
45	3
46	19q13.3
47	1
48	1
49	1
50	1
51	X
52	22q13.33
53	19
54	19
55	6
56	3

586
TABLE 6

SEQ_ID	GENOMIC_LOCATION
57	4
58	13
59	11
60	1
61	19
62	13
63	12
64	6
65	1
66	4
67	7
68	20
69	11
70	12p13
71	1p32.1-33
72	3
73	14
74	14
75	6p11.2-12.3
76	11q
77	15
78	19
79	2
80	9
81	9
82	9
83	11
84	2p13
85	11
86	1p36.2-36.33
87	1
88	17
89	20p13
90	7q22-q31.1
91	4
92	7
93	17q25
94	11q13.3
95	8q22-q23
96	11
98	8
99	7
100	4
101	2
102	6q16.3-22.1
103	4
104	5
105	5
106	Xp11.3
107	12pter-p13.31
108	4
109	19
110	10cen-q26.11
111	19p13
112	17

587
TABLE 6

SEQ_ID	GENOMIC_LOCATION
113	17
114	17
115	6
116	14
117	14
118	10
119	3
120	1
121	10
122	14
123	14
124	4
125	15q21.3
126	15
127	6
131	17
132	11q
133	11q
135	1p36.12
136	16
137	18
138	3
139	2
140	11q13
141	7q33-q35
142	7q33-q35
143	4
144	6q25.3-27
145	8
146	8
147	17
148	17
149	11
150	22
151	17
152	17
153	Xq22
154	5
155	14
156	13q12.11-12.3
157	9
158	22q12
159	22q12
160	10
161	10
163	15q15
164	15q15
165	17
166	4
167	4p16
168	4p16
169	4p16
170	11
171	5
172	14

588

TABLE 6

SEQ ID	GENOMIC LOCATION
173	15
174	15
175	8p21.3-q11.1
176	8p21.3-q11.1
177	3p21-p12
178	19
179	16
180	22q12
181	X
182	6
183	5p13
184	5p13
186	12q
187	10
188	3
189	2
190	22
191	7
192	19
193	14
194	11q22
195	11q22
196	11q22
197	15
198	4
199	4q28-q32
200	6p21.1-22.2.
201	12
202	2p24
203	2p24
204	1
205	14
206	18
207	20
208	19
209	1
210	4
211	1
212	4
213	3
214	3
215	9
216	1
217	11
218	7q
219	7q
220	17
221	5
222	15
223	10
224	8
225	8
226	8
227	9
228	8

589

TABLE 6

SEQ_ID	GENOMIC_LOCATION
229	18
230	7pter-p22
231	7
232	17
233	17
234	17
235	10
236	1p34.1-1p35.
237	4p16-p15
238	4p16-p15
239	4p16-p15
240	4p16-p15
241	4p16-p15
242	20
243	6p12.3-21.2
244	2
245	6p21.2-22.1
246	19q13.4
247	11q13.3
248	6
249	18
251	3
252	11
253	11
254	19
255	6
256	2
257	Xp22
258	Xp22
259	20q12-13.1
260	20q12-13.1
261	20q12-13.1
262	20q12-13.1
263	16
264	4
265	11
266	9q34.2-34.3
267	20
268	12
269	7
270	9
271	20
272	14
273	9p34.1-35.1
274	17
275	10
276	10
277	10
278	10
279	1p36.11-36.31
281	11
282	3p21.3
283	14
284	1
285	1

590

TABLE 6

SEQ ID	GENOMIC LOCATION
286	11
287	22q12.3-13.2.
288	11
289	17q21.3
290	2
291	1p22.2-31.1
292	1p22.2-31.1
293	19
294	Xq23
295	Xq23
296	10q25.3-q26.2
297	1
298	20q12-13.2
299	6
300	2
301	2
302	12q24.1
303	20q11.21-13.13
304	8
305	1
306	12q
307	18q12-q21
308	14
309	5
310	2
311	2q32.1-q36.3
312	12p13.32
313	5
314	13
315	22q13.1
316	18
317	11
318	11
319	3
320	2
321	4q27
322	11
323	11
324	Xq12.1-13.
325	8
326	22.
328	3
329	9
330	9
331	5
332	22q13.31-13.33
333	22q13.31-13.33
334	10
335	3
336	15
337	5
338	5
339	3p
340	3
341	3

591
TABLE 6

SEQ_ID	GENOMIC_LOCATION
342	5
343	12
344	14
345	14
346	3
347	16
348	4
349	6q24.3-25.3
350	5
351	1
352	9q31.1-31.3
353	2
354	16
355	6
356	5q13
357	17
359	6
360	19
361	5
362	11q14.3-q21
363	12q
364	22q13.31-13.33
365	9
366	2
367	2
368	22q11.1-q11.2
369	16
370	17
371	11
372	20
373	20
374	20
375	20
376	15q13-q14
377	1
378	16p13.3
379	20
380	3
381	1p32-p31
382	q42.2-43
383	X
384	15
385	16
386	8
387	3p25-p24
388	11
389	7
390	7
391	2
392	22
393	22
394	12
395	12
396	12
397	12

592
TABLE 6

SEQ ID	GENOMIC LOCATION
398	12
399	14
400	2
401	3
402	1
403	19q13.4
404	19q13.4
405	17q12-q21
406	17q12-q21
407	17q12-q21
408	10
409	10
410	12p13.3
411	Xq13.2-21.1
412	12p13
413	12p13
414	19
415	7q22
416	19
417	8
418	8
419	18
420	7
421	1q25.1-31.3
422	8
423	11
424	11
425	16
426	19p13.3
427	11
428	2
429	7
430	8
431	4
432	4
433	19
434	19
435	5
436	14
437	3
438	22q13
439	20p11.22-12.2
440	2
441	2
442	4
443	1
444	1
445	8
446	8
447	7p11.2-q11.2
448	8
449	19
450	6
451	7q22-q31.1
452	19

593

TABLE 6

SEQ ID	GENOMIC LOCATION
453	4q22-q24
454	17
455	8
456	1
457	1
458	19
459	7q33-q35
460	7q33-q35
461	14
462	4
463	2
464	17
465	17
466	19
467	12
468	11q22
469	11
470	10cen-q26.11
471	22q12-13.
472	20q13.33
473	1
474	1
475	17
476	18
477	3
478	12
479	16
480	5
481	2
482	2q33-q34
483	19
484	4
485	12p12.3-13.2
486	3
487	3
488	19
489	19
490	19
491	10p12
492	17
493	10
494	12
495	18
496	13
497	10
498	16
499	22q12.2
500	X
501	3
502	15
503	3
504	19
505	19
506	19
507	5

594

TABLE 6

SEQ ID	GENOMIC LOCATION
508	7q21.2-q31.1
509	2
510	2p13
511	1p21-p13
512	7
513	11
514	16
515	9
516	18
517	10
518	5
519	10
520	21q22.3
521	7
522	3p21.1-p14.2
523	1q21
525	1q25.1-31.1
526	7q35
527	9
528	1
529	5q32
530	19
531	2
532	18
533	1
534	22
535	13
536	X
537	4q21-q25
538	1q23-q25.1
539	18
540	22q12
541	3p24
542	19p13.1
543	2
544	14
545	6p21.1-21.3
546	12
547	22q12-13.
548	22q12-13.
549	22q12-13.
550	17
551	14
552	15
553	15
554	1
555	1q23.2-24.3
556	15q21.1
557	3
558	12q24.1
559	17
560	19
561	19
562	9
563	5

595
TABLE 6

SEQ_ID	GENOMIC_LOCATION
564	7
565	2
566	11
567	5
568	5
569	1
570	3
571	14q24.3
572	11q22.3-q23.1
573	11q22.3-q23.1
574	17
575	4
576	12q
577	13q14.2-14.3
578	3
579	10
580	15
581	10
582	10
583	6
584	11
585	X
586	17
587	7
588	3
589	2
590	8
591	20
592	17
593	17
594	5
595	1
596	16
597	7
598	7
599	7
600	17
601	2
602	9
603	7
604	19
605	16q22
606	17
607	22
608	11q
609	11q
610	9
611	20
612	2
613	6
614	6
615	15
616	22q13.32
617	11
618	11

596

TABLE 6

SEQ ID	GENOMIC LOCATION
619	15
620	15q21.3
621	15q21.3
622	14
623	14q
624	3
625	3
626	2p22-2p21
627	7
628	1
629	06
630	X
631	8
632	8
633	17q25.2-q25.3
634	16
635	13
636	13
637	1
638	19
639	19
640	9
641	19q13.2
642	19q13.2
643	1p36.1
644	19
645	1
646	Xq23
649	2q35
650	17
651	7
652	1q21
653	4
654	14
655	2
656	1p36.2-p35
657	2
658	16
659	3
660	6
661	10
662	10
663	16
664	3
665	3
666	17
667	12
668	16p11.2
669	11
670	17
671	2
672	12p13.3
673	12p13.3
674	5
675	2

597
TABLE 6

SEQ ID	GENOMIC LOCATION
676	6p21.2-21.31
677	19
678	19
679	19
680	3
681	2q14
682	12
683	7

598
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
1967	A	75	509	DPKAQLPEPLRVLWTAHLVAMAPGSRTSLLLAFALLC LPWLQEAGAVQTVPLSRLFDHAMLQAHRAHQLAIDTY QEFEETYIPKDQKYSFLHDSQTSFCFSDSIPTPSNME ETQOKSNLELLRISLLLIESWLEPVRILMSIVPN
1968	A	75	509	DPKAQLPEPLRVLWTAHLVAMAPGSRTSLLLAFALLC LPWLQEAGAVQTVPLSRLFDHAMLQAHRAHQLAIDTY QEFEETYIPKDQKYSFLHDSQTSFCFSDSIPTPSNME ETQOKSNLELLRISLLLIESWLEPVRILMSIVPN
1969	A	75	509	DPKAQLPEPLRVLWTAHLVAMAPGSRTSLLLAFALLC LPWLQEAGAVQTVPLSRLFDHAMLQAHRAHQLAIDTY QEFEETYIPKDQKYSFLHDSQTSFCFSDSIPTPSNME ETQOKSNLELLRISLLLIESWLEPVRILMSIVPN
1970	A	75	509	DPKAQLPEPLRVLWTAHLVAMAPGSRTSLLLAFALLC LPWLQEAGAVQTVPLSRLFDHAMLQAHRAHQLAIDTY QEFEETYIPKDQKYSFLHDSQTSFCFSDSIPTPSNME ETQOKSNLELLRISLLLIESWLEPVRILMSIVPN
1971	A	1764	403	KAAKKALCWLEPPQCAGLEGLGWVWSCSVSTGPRMQA LVLLLCIGALLGHSSCQNPASPPEEGSPDPDSTGALV EEEDPFFKVPVNKLAAAVSNFGYDLRVRSSMSPTTN VLLSPLSVATALSALSLGAEQRTESIHRALYYDLIS SPDIHGTYKELLDTVTAPQKNLKSASRIVFEKKLRIK SSFVAPLEKSYGTRPRVLTGNPRLDLQEINNWWQAQM KGKLARSTKEIPDEISILLG\VAHFKGQ\WETKFDS RKTSLEDFYLDERTVRVPMMSDPKAVLRYGLSDLS CKIAQLPLTGSMSTIFFLPLKVTQNLTLIEESLTSEF IHDIDRELKTVQAVLTVPKLKLSEGEVTKSLQEMKL QSLFDSPDFSKIITGKPIKLTQVEHRAGFEWNEDGAGT TPSPGLQPAHLTFPLDYHLNQPFI FVLRD TDTGALLF IGKILDPRGP
1972	A	3	147	QPLNHYFICSSHNTYLVGDQLCGQSSVEGYIRCSGGR EGVQLMRGTM
1973	A	2	2117	FVVAASGGCWFVLGERRAGSLLSASYGTFAMPGMVLF GRRWAIASDDLVPFGFFELVVRVLWWIGILTLYLMHR GKLDCAGGALLSSYLIVLMILLAVVICTVSAIMCVSM RGTICNPGPRKSMKLLYIRLALFFPEMVWASLGAAW VADGVQCDRTVVNGIIATVVVSWIIIAATVVSIIIVF DPLGGKMAPYSSAGPSHLDSDSSQLLNGLKTAATSV WETRIKLLCCCIGKDDHTRVAFSSTAELFSTYFSDTD LVPSDIAAGLALLHQQODNIRNNQEPAQVVCHAPGSS QEADLDAELKNCHHYMQFAAAAYGWPLYIYRNPLTGL CRIGGDCCRSKNPQMT/MVGGDQLQL/CTSAPILHT HRAAVQGLHPRQLPWTRFTELPFLVALDHRKESVVA VRGTMSLQDVLTDLSAESEVLDVECEVQDRLAHKGIS QAARYVYQRLINDGILSQAFSIAPEYRLVIVGHSLGG GAAALLATMVRAAYPQVRCYAFSPPRGLWSKALQEYS QSFIVSLVLGKDVI PRLSVTNLEDLKRRI LRVAHCN KPKYKILLHGLWYELFGGNPNNLPTELDGGDQEVLTQ PLLGEQSLLTRWSPAYSFSSDSPLDSSPKYPPLYPG RIIHLQEEGASGRFGCCSAAHYSAKWSHEAEFSKILI GPKMLTDHMPDILMRALDSVSDRAACVSCPAQGVSS

599
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VDVA
1974	A	62	616	EVHQGTEVRDSEVRRRPQARGPLMPAERAGRQRWLVP ALQPRRGGLRR*RGAVRQHGAHPHGLLLQDQKI PALP GRKQAGSLHAPGTEGEPDHGGDPVLDAGIQHHRQQRH PTADHLNPGEHRRGEAHVRAAV*PAAGAEGA AKERRA HQANTALQVHRR*LGSFAELRLLLRKPGRTSVWPSM
1975	A	337	440	PLALCLAPAASLHELCAAKVSEVLHNRVHRTEEV
1976	A	1454	1101	AFYNANSCLNVFCFCFCFWRQSRCISQAGVQWCDLSS LQPPLPRFKRFSYLSLPSSWDYRHAPPCPANFCI\LV ETGLCHIGHACLELLTSGDPPALASQSAGITGMSHST QPCIAVS
1977	A	2	1454	DDFVGVL SATAQVCTMAARLVSRGAVRAAPHSGPL/ AVLAQVRR\STD TVYDVVVSGGGLVGAAMACALGYD IHFHDKKILLLEAGPKKVLEKLSETYSNRVSSISPGS ATLLSSFGAWDHICNMRYRAFRMQVWDACSEALIMF DKDNLDDMGYIL\ENDV\IMHAFTKQLEAVSDRVTVL YRSKAIRYTWPCPFPMADSSPWVHITLGDGSTFQTKL LIGADGHNSGVRQAVGIQNVSWNYDQSAVVATLHLSE ATENNVAVQRFPLPSGP IALLPLSDTLSSLVWSTSHEH AAELVSMDEEKFVDAVNSAFWSDADHTDFIDTAGAML QYAVSLLKPTKVSARQLPPSVARVDAKSRVLFPLGLG HAAEYVRPRVALIGDAHRVHPLAGQGVNMFGD ISS LAHHLSTA AFNGKDLGSM SHLTGYETERQRHNTALLA ATDLLKRLYSTSASPLVLLRTWGLQATNAVSPLEQI MAFASK
1978	A	3692	3395	LKDSLRLRFFFFEMESCSVTRLECSGVISAHNRLRLPG SSNSPTSASQVAGTTGMHPHTQLIFVFS AETGFPHAG QDGLDLL/NLVISPPWPPKVLGLQA
1979	A	65	265	SALLGLPSSWDYRRPPPRPANFLYF**RRGFTVLARM VSIC*PRDPPASASRSAGISGVSRGRPPS
1980	A	751	176	LPGADYGGGHL SLRLFHL LLSAAWVPDESQVTLSA ICVLSTV LIMEFPDLGKHCSEKTCQLDLFPVKCDAC KQDFCKDHF PYAAHKCPFAFQKDVHVPVCPLCNTPI P VKKGQIPDVVGDHIDRDCDSHPGKKKEKIFTYRCSK EGCKKEM LQMVCAQCHGNFCIQHRHPLDHSCRHGS PTIKAG
1981	A	250	118	DSLTRLPALCSLQLGRKVETITIIYDCEGLGLKHLWK PAVEAYG
1982	A	235	1157	SIQEKCFDSSCGRNSLLSFSLSYKESHKTFIFYCWVY RLCIWI\TAIWQYESLKS RVQSYFDGIKADWLDSIRP QKEGDFRKEINKWNNLSDGQRTVTGIIAANVLVFC WRVPSLQRTMIRYFTSNPASKVLCSPMLLSTFSHFSL FHMAANMYVLWSFSSSIVNILGQEQFMAVYLSAGVIS NFVSYLGKVATGRYGPSLGASGAIMTVLAAVCTKIPE GRLAII FLPMFTFTAGNALKAI IAMDTAGMILGWKFF DHAHLGGALFGI WYVTYGHELIWKNREPLVKIWHEI RTNGPKKGGGSK
1983	A	289	392	RAFAEAMRGYHGDRGSHPRPARFADQQHMDVGP A
1984	A	98	1474	MAWASRLGLLLALLLPVVGASTPGTVVRLNKAALSYV SEIGKAPLQRALQVTVPHFLDWSGEALQPTRIRILNV

600
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				HVPRLLHLKFIAGFGVRLLAANFTFKVFRAPEPELELT LPVELLADTRVTQSSIRTPVVSISACSLFSGHANEFD GSNSTSHALLVLVQKHIKAVLSNKLCLSI SNLVQGVN VHLGTLIGLNPVGPESQIRYSMVSVPTVTSDYISLEV NAVLELLGKPIILPTDATPFVLPRHVGTEGSMATVGL SQQLFDSALLLLQKAGALNLDITGQLRSDDNLLNTSA LGRLIPEVARQFPEPMPVVLKVR LGATPVAMLHTNNA TLRLQPFVEVLATASNSAFQSLFSLDVVNLRLQLSV SKVKLQGTTSVLGDVQLTVASSNVGFIDTDQVRTLMG TVFEKPLLDHLNALLAMGIALPGVNLHYVAPEIFVY EGYVVISSGLEFYQS*
1985	A	541	176	GPHTSNRPRXRHCTXGPSTXXTXAGSGYSPA HGRAWG APCXSX* RSPGPRGGRESGTCRPAAAPAPAPAGGCRA GTGA WPPGSATSPRC* SPAAPRGAGPQPGSGGSHGGT ARMCACKLAAS
1986	A	2390	1943	AGRRLTQAGTLLGTALAFGTRLLVSSDMKSWSTVLAV MGKAFSEAAFTTAYLFTSELYPTVLRQTGMGLTALVG RLGGS LAPLAALLDGWLSL PKLTYGGIALLAAGTAL LLPETRQAQLPETIQDVERKSAPTSLQEEEMP MKQVQ N
1987	A	1	555	KKVGNYYTTP IYRFRMKCHLCVNYIEMQTD PANC DYV IVSGAQRKEERWDMADNEQVLT TGERHPLTCLGAL/D PESALGPPKPSRALI VAEHEKKQKLETDAMFRLEHGE ADRSTLKKALPTLSHIQEAQSAWKDDFALNSMLRRRF RVRGAPARGQRCMVDQGP GPALPPPHPSFEQATCTF
1988	A	2867	847	GLPGIPGLPGFPVAGPPGITGFPFGFIGSRGDKGAPG RAGLYGEIGATGDFGDIGDTINLPGRPGLKGERGTTG IPGLKGFFGEKGTEGDIGFPGITGVTGVQGP PGLKGQ TGFPGLTGPPGSQGELGRIGLPGGKGDDGWPGAPGLP GFPGLRGIRGLHGLPGTKGFPGPSGSDIHGDPGFPGP PGERGDPGEANTLPGPVGVPGQKGDQGAPGERGPPGS PGLQGFPGITPPSNISGAPGDKGAPGIFGLKGYRGPP GPPGSAALPGSKGDTGNPGAPGTPGTKGWAGDSGPQG RPGVFGLPGEKGPGEQGFMGNTGPTGAVGDRGPKGP KGDPGFPGAPGTVGAPGIAGIPQKIAVQPGTVGPQGR RGPPGAPGEMGPQGPGEPPGFRGAPGKAGPQGRGGVS AVPGFRGDEGPIGHQGP I GQEGAPGRPGSPGLPGMPG RSVSI GYLLVKHSQTDQEP MCPVGMNKLWSGYSLLYF EGQEKAHNQDLGLAGSCLARFSTMPFLYCNPGDVCYY ASRNDKSYWLSTTAPLPMMPVAEDEIKPYISRCSVCE APAIAIAVHSQDVSI PHCPAGWRSLWIGYSFLMHTAA GDEGGGQSLVSPGSCLEDFRATPFIECNGGRGTCHYY ANKYSFWLT TIPEQS FQGS PSADTLKAGLIRTHISRC QVCMKNL
1989	A	1	777	LIYNEDMICWIESRESSNQLKCIQITKAGGLTDEWTI NILQSFHNVQQMAIDWLTRNLYFVDHVGDRIFVCNSN GSVCVTLIDLELHNPKAIAVDPIAGKLFFTDYGNVAK VERCDMDGMNRTRI IDSKTEQPAALALDLVNKL VYVW DLYLDYVGVVDYQGNRHAVIQGRQVRHLYGITVFED YLYATNSDSYNIVRISR FNGTDIHS LIKIENAWGIRI

601
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				YQKRTQPTVRSHACEVDPYGMPPGGCSHICLLSSSYTK
1990	A	1	777	LIYNEDMICWIESRESSNQLKCIQITKAGGLTDEWTI NILQSFHNVQQMAIDWLTRNLYFVDHVGDRIFVCNSN GSVCVTLIDLELHNPKAIAVDPIAGKLFFTDYGNVAK VERCDMDGMNRTRIIDSKEQPAALALDLVNKLIVYVW DLYLDYVGVVDYQGKNRHAVIQGRQVRHLYGITVFED YLYATNSDSYNIVRISRFNGTDIHSLIKIENAWGIRI YQKRTQPTVRSHACEVDPYGMPPGGCSHICLLSSSYTK
1991	A	1620	1214	LPFLSFFLSFFLFFLRWSFALIAQAGVQWCNFGSPQP PPPGEKRFSCLSLLSSWDYRHTPPCLANSVFLVDTGF LHVQAGLELPTSGDPPTSASQSAGITSVSHCAQPVT AISKEEREQAEGPDSQGTGSSAGQ
1992	A	1	660	GFHPNTTHYRARAARAGAGSFVGEVSAVDKDFGPNG EVRYSFEMVQPDFELHAISGEITNTHQFDRESLMRRR GTAVFSFTVIATDQGI PQPLKDQATVHVYMKDINDNA PKFLKDFYQATISESAANLTQVLRVSASDVDEGNNG IHYSI IKGNEERQFAIDSTSGQVTLLIGKLDYEATPAY SLVIQAVDSGTIPLNSTCTLNIDILDENDNTPFFP
1993	A	1	660	GFHPNTTHYRARAARAGAGSFVGEVSAVDKDFGPNG EVRYSFEMVQPDFELHAISGEITNTHQFDRESLMRRR GTAVFSFTVIATDQGI PQPLKDQATVHVYMKDINDNA PKFLKDFYQATISESAANLTQVLRVSASDVDEGNNG IHYSI IKGNEERQFAIDSTSGQVTLLIGKLDYEATPAY SLVIQAVDSGTIPLNSTCTLNIDILDENDNTPFFP
1994	A	2	271	GSVALHVEKLPNEPNRLILHGFLENVHFFHTNFLV SQLIRAGKPYQLQVALPPVSPQIYPNERHSIRCPESG EHYEVTLLHFLQEYL
1995	A	289	418	LWTLYRHKQQVQHNHNSNRLSCRPSQEDRATHTIMVLD KENTLS
1996	A	3	673	RNFRVDDFVAELKLKQVRWTPAAP*SKETTQGLRRLH VNGRCEPKGLDPEMGRSSDTEESRSKRKKKHRRRS SSSSSSDSRTYSRKKGGRKSRKSRWSRDLQPRSHS YDRRRRHRSSSSSYGSRRKRSRSGRGKSYRVQR SRSKSRTRRSRSPRLRSHSRSSERSSHRTRSRSRD RERRKGRDKEKREKEKDKGDKELHNI KRGESGNKA GLE/HSATS*TGQSQT TAGS*SCCKS**SIESQRKK* GRSKE/QERRKTKPPW*NK*KE*KFGGRRRRPDLKKR LRDCGACTSTGGVSPKVWTQKWDVGHQILKKKAEARE KRNTVDGPPRAVLQIVEHTAERKEEGNQDQSQDLGPE IFSLVHILMIEDAGIDQAVALLMAPEGNEVEVVQGV GNPIEFRGLGQKAEQEGPGQDLVSVLIVVAVKGPVTE ERVVGLGIENDVRAEIKRKEKRRRIKGRTRNYITSNV GNLETSKLD
1997	A	279	762	VGNFQRQLAEAKEDNCKVTIMLENVLASHSKMQGALE KVQIELGRDSEIAGLKKERDLNQQRVQKLEAEVDQW QARMLVMEDQHNSIEISLQKALGVAREDNKRKLAMSLE QALQTNHLQTKLDHIQEQLSKELERQNLETFKDRM TEESKVEAELHAE
1998	A	3	1434	PPNMDNSMGTEETITVLKGSSTSMACITDGTAPSPMAW LRDGQPLGLDAHLTVSTHGMVLQLLKAETEDSGKYTC

602
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				IASNEAGEVSKHFILKVLVPPSFQKLWEIGNMLDTGR NGEAKDVIINNPIISLYCETNAAPPPTLTWYKDGHP SSDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGE DSLQYDVRVLVPPIIKGANSDLPEEVTVLVNKSALIE CLSSGSPAPRNSWQKDGQPLLEDDHHKFLSNGRILQI LNTQITDIGRYVCVAENTAGSAKKYFNLNVHVPSPVI GPKSENLTVVVNNFISLTCEVSGFPPDLSWLKNEQP IPLNTNTLIAPGGRTLQIIRAKVSDGGEYTCIAINQA GESKKKFSLTVYVPPSIKDHDSESLSVNVREGTSVS LECESNAVPPPVTWYKNGRMITESTHVEILADGQML HIKKAEVSDTGQYVCRAINVAGRDDKNFHLNVY
1999	A	2	1333	RSGEGFHVNSS*TWVSRS*EMDETPGSEVPGDKAAEE QGDDQDSEKSKPAGSDGERRGVKRQRDEKDEHGRAY EFREEAYHSRSKSPLPPEEEAKDEEEDQTLVNLDTYT SDLHFQVSKDRYGGQPLFSEKFPTLWSGARSTYGVTK GKVCFEAKVTQNLPMKEGCTEVSLLRVGSVDFSRPQ LGEDEFSYGFDDGRGLKAENGQFEFFGQTFGENDVIGC FANFETEEVELSFSKNGEDLGVAFWISKDSLADRALL PHVLCNKCVVELNFGQKEEPPFPPEEFVFIHAVPVE ERVRTAVPPKTIEECEVILMVGLPGSGKTQWALKYAK ENPEKRYNVLGAETVLNQMRMKGLEEPEMDPKSRDLL VQQASQCLSKLVQIASRTKRNFILDQCNVYNSGQRRK LLLFKTFSRKVVVVVPNEDDWKKRLELRKEVEGRVFP
2000	A	1	1060	IIFLFF*PYLQSVIFLFVIRGLEMKYGNIEIMNKDPVF RISPRSRETHPNPEEPEEEDVQAEVQAANALTAP NLEEEFVITASCLHKEYYETKKSCFSTRKKKIAIRNV SFCVKKGEVLGGLGHNGAGKSTSIKMITGCTVPTAGV VVLQGNRASVRQQRDNSLK/FLGYCPQENSLWPKLTM KEHLELYAAVKGLGKDAALSIS*LVEALKLQEQLKAP VKTLSEGIKRKLCFVLSILGNPSVLLDELFTGMDPE GQQQMWOILQATIKNQERGALLTTHYMSEAKSLCDRV AIMVSGTLRCIGSIQQL/KKFGKDYLLEIKMKEPTQV EALHTEILKLFPPQAAWQERYSSL
2001	A	1	2543	TISSSPKWRLSGWRAPCCWGFVAGGPGDPFPAAEA LEDESGTLLRSGGGAGEQWQQGLRWRPRSGMCESYSR SLLRVSAQICQALGWDSVQLSACHLLTDVLQRYLQQ LGRGCHRYSELYGRTPILDDVGEAFQLMGVSLHELE DYIHNIETPVTFPHQIPSPFVSKNNVLQFPQPGSKDAE ERKEYIPDYLPIVSSQEEEEEEQVPTDGGTSAEAMQ VPLEEDDELEEEIINDENFLGKRPLDSPEAEELPAM KRPRLLSTKGDTLDVVLEAREPLSSINTQKIPMPLS PVHVQDSTDLAPSPPEPPMLAPVAKSQMPTAKPLETK SFTPKTKTKTSSPGQKTKSPKTAQSPAMVGSPIRSPK TVSKEKKSPGRSKSPKSPKSPKVTTHIPQTPVRPETP NRTPSATLSEKISKETIQVKQIQTPPDAGKLNSNQ KKAVVADKTIEASIDAVIARACAEREPPFEFSSGSE SEGDIFTSPKRISGPECTTPKASTSANSFTKSGSTPL PLSGGTSSSDNSWTMDASIDEVVRKAKLGTSPSNMPPN FPYISSPSVSPPTPEPLHKVYEEKTKLPSSVEVKKKL KKELKTKMKKKEKQDREREKDKNKDKSKEKDKVKEK

603
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EKDKETGRETKYPWKEFLKEEEADPYKFKEFEDVD PKVKLKDGLVRKEKEKHKDKKKDREKGGKDKDKREKE KVKDKGREDKMKAPAPPLVLPPELALPLFSPATASR VPAMLPSLLPVLPEKLFEEKEKPKEKEKKKDKKEKKK KKEKEKEKKEKEREKEKREKREKEKEKHKHEKIKV EPVALAPSPVIPRLTLRVGAGPDKIRRRRAGAH
2002	A	2	1736	QNSVSDKWGKPLVIDKLEMAKVEGLWNLFPAVSG LSHVDYALIAEETGKCFAPDVFNCQAPDTGNMEVLH LYGSEEQKKQWLEPLLQGNITSCFCMTEPDVASSDAT NIECSIQRDEDSYVINGKKWWSSGAGNPKCKIAIVLG RTQNTSLSR*LNNSD*ETCVGMSQSSSYLGNLLKIHC LDSQIIM*DMRVNVIYLYFTSIF*QVFLENIIGSIAE HSSLWNFQY*KVLLNYQSCLD*IIRQIFSDLCNEVIR CLDQRQ*S*NV*LYI*VPSYHC*AVRSFNQTTHLFSN HCFCSRSQPASDYVGVRLHSSHSHHCLHDYMKTSK RQLGFCLLSVLFFFLANFF*YNFSFD*\HKQHSMILV PMNTPGVKIIRPLSVFGYTDNFHGGHFEIHFNQVRVP ATNLILGEGRGFEISQGRLGPGRIHHCMRTVGLAERA LQIMCERATQRIAFKKKLYAHEVVAHWIAESRIAIEK IRLLTLKAAHSMDTLGSAGAKKEIAMIKVAAPRAVSK IVDWAIQVCGGAGVSQDYPLANMYAITRVLRLADGPD EVHLSAIAATMELRDQAKRLTAKI
2003	A	2240	506	RRPPEGSGGGRRTRARMPLPWSLALPLLLSWVAGGF GNAASARHHGLLASARQPGVCHYGTKLACCYGWRNNS KGVCEATCEPGCKFGCEVGNPKCRCFPGYTGKTCSDQ VNECGMKPRPCQHRVCNTHGSYKCFCLSGHMLMPDAT CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN GRDCLDIDECASGKVICPYNRRVCNTEFGSYCKCHIG FELQYISGRYDCIDINECTMDSHTCSHHANCFNTQGS FKCKCKQGYKGNLRCSAIPENSVKEVLRAPGTIKDR IKKLLAHKNSMKKKAKIKNVTPEPTRTPKVNLPQPF NYEEIVSRGGNSHGG\KKGNEEKMKEGLEDEKREEKA LKD*HRRERPFRG\DVFFPKVNEAGEFGLIL\VQRKA LTSKLEHKADLNI SVDCSFNHG\ICDW\KQDR\EDDF DW\NPADR\DNAI\GFY\MAVPGLWQGHK\KDIGRLK LLLPLDLPQSNFCLLFDYRLAGDKVGKLRVFKNSNN ALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEAEER GKGKTGEIAVDGVLLVSGLCPDLSLLSVDD
2004	A	2	469	KGTKNQGFNYPWDVAVNSEGKILVSDTRNHRIQLFGP DGVFLNKYGFEGALWKHFDSPRGVAFNHEGHLVVTDF NNHRLLVHPDCQSARFLGSEGTGNGQFLRPQGVAVD QEGRIIVADSRNHRVQMFESNGSFLCKFGAQSGFGQ MDRPSGIA
2005	A	4135	639	QCGPEAASAGSCSAETPSPPPRAPGRGPIMFSRKKRE LMKTPSISKKNRAGSPSPQPSGELPRKDGADAVFPGP SLEPPAGSSGVKATGTLKRPTSLSRHASAAGFPLSGA ASWTLGRSHRSPLTAASPGELPTEGAGPDVVEDISHL LADVARFAEGLEKLKECVLHDDLLEARRPRAHECLGE ALRVMHQIIISKYPLLNTVETLTAAGTLIAKVKAHYE SNNDLEKQEFKALETIAVAFSSTVSEFLMGEVDSST

604
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LLAVPPGDSSQSMESELYGPGSEGTPPSLED CDAGCLP AEEVDVLLQRCCEGGVDAALLYAKNMAKYMKDLSYLE KRTTLEMEFAKGLQKIAHNCRQSVMQEPHMPLLSIYS LALEQDLEFGHSMVQAVGTLQTQTFMQPLTLRRLEHE KRRKEIKEAWHRAQRKLQEAESNLRKAKQGYVQRCED HDKARFLVAKAEEEQAGSAPGAGSTATKTLDKRRRLE EEAKNKAEEAMATYRTCVA DA KTQKQELEDTKVTALR QIQEVIRQSDQTIKSATISYYQMMHMQTAPLPVHFQM LCESSKLYDPGQQYASHVRQLQDQEPDVHYDFEPHV SANAWSPVMRARKSSFNVSDVARPEAAGSPPEEGGCT EGTPAKDHRAGRGHQVHKSWPLSISDSDSGLDPGPGA GDFKKFERTSSSGTMSSTEELVDPDGGAGASAFEQAD LNGMTPELPVAVPSGPFRRHEGLSKAARTHRLR\KLRT PAKCRECNSYVYFQGAEECECLACHKKCLETLAIQC GHKKLQGRLLQLFQDFSHAARSAPDGVFIVKKCVCE IERRALRTKGIYRVNGVKTRVEKLCQAFENGKELVEL SQASPHDISNVLKLYLRQLPEPLISFRLYHELVLGAK DSLKAEAEAKAASRGRODGESESEAVAVALAGRLRELL RDLPPENRASLQYLLRHLRRIVEVEQDNKMTPGNLGI VFGPTLLRPRPTEATVSLSSLVDYPHQARVIETLIVH YGLVFEEPEETPGGQDESSNQRAEVVVQVPYLEAGE AVVYPLQEAAADGCRESRVVSNDSDSDLEEASELLSS SEASALGHLSFLEQQQSEASLEVASGSHSGSEEQLEA TAREDGDGEDGPAQQLSGFNTNQSNVQLQAPLPPMR LRGGRMTLGSCRERQPEFV
2006	A	3	628	SVGALDTFIAAVYEHAVILPNRAETPVSKEEALLLMN KNIDVLEKAVKLAQGAHIIVTPEDGIYGWIFTRES IYPYLEDIPDPGVNWI PCRDPWRNH*NIVSLRKCLLN \RFGNTPVQQRLSCLAKDNSIYVVANIGDKKPCNASD SQCPPDGRYQYNTDVVFDSQGKLLARYHKYNLFAPEI QFDFPKDSELVTFTDTPFGKIGIIT
2007	A	1375	1453	RTFTS*CSVSCGRGVQQRHVGCQIGTHKIARETECNP YTRPESERDCQGPRCPLYTWRAEEWQEVSRATKGYLP GISRVRLPSSHLFPKPEKSPSTVTMLALSQKVHCQ TRAFAPTRVGELLVFKQFL
2008	A	2679	1435	LLSTYIKFINLFPETKATIQQVLRAGSQLRNADVELQ QRAVEYLTLSVASTDVLATVLEEMPPFPERESSILA KLKRKKGPGAGSALDDGRRDPSSNDINGGMEPTPSTV STPSPSADLLGLRAAPPPAAPPASAGAGNLLVDVFDG PAAQPSLGPTPEEAFLSPGPEDIGPPIPEADELLNKF VCKNNGVLFENQLLQIGVKSEFRQNLGRMYLFYGNKT SVQFQNFSPTVVHPGDLQTQLAVQTKRVAAQVDGGAQ VQQVLNIECLRDFTPLLSVRFRYGGAPQALTLKLP VTINKFFQPTMAAQDFFQRWKQLSLPQQEAQKIFKA NHPMDAEVTKAKLLGFGSALLDNVDNPNENFVGAGII QTKALQVGCLLRLEPNAAQMYRLTLRTSKEPVSRHL CELLAQOF
2009	A	153	1994	MGALRPTLLPPSLPLLLLLMLGMGCWAREVLVPEGPL YRVAGTAVSISCNVTGYEGPAQQNFQFLYRPEAPDT ALGIYSTKDTQFSYAVFKSRVVAGEVQVQRLQGDVV

605
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LKIARLQAQDAGIYECHTPSTDTRYLGSYSGKVELRV LPDVLQVSAAPPGPRGRQAPTSPPRMTVHEGQELALG CLARTSTQKHTHLAVSFGRSVPEAPVGRSTLQEVVGI RSDLAVEAGAPYAERLAAGELRLGKEGTDTRYRMVVG AQAGDAGTYHCTAAEWIQDPDGSWAQIAEKRAVLAVH DVQTLSSQLAVTVGPGERRIGPGEPELLELCNVSGALP PAGRHAAYSVGWEMAPAGAPGPGRLVAQLDTEGVGSL GPGYEGRHIAMEKVASRTYRLRLEAARPGDAGTYRCL AKAYVRGSGTRLREAASARSRPLPVHVREEGVVLEAV AWLAGGTVYRGETASLLCNISVRGGPPGLRLAASWWV ERPEDGELESSVPAQLVGGVGQDGVAEELGVRPGGGPVS VELVGPRSHRLRLHSLGPEDEGVYHCAPSAWVQHADY SWYQAGSARSGPVTVPYPMHALDTLFLVPLLVTGTGVAL VTGATVLGTITCCFMKRLRKR*
2010	A	153	1994	MGALRPTLLPPSLPLLLLLMLGMGCWAREVLVPEGPL YRVAGTAVSISCNVTGYEGPAQQNFEWFLYRPEAPDT ALGIVSTKDTQFSYAVFKSRVVAGEVQVQRLQGDVV LKIARLQAQDAGIYECHTPSTDTRYLGSYSGKVELRV LPDVLQVSAAPPGPRGRQAPTSPPRMTVHEGQELALG CLARTSTQKHTHLAVSFGRSVPEAPVGRSTLQEVVGI RSDLAVEAGAPYAERLAAGELRLGKEGTDTRYRMVVG AQAGDAGTYHCTAAEWIQDPDGSWAQIAEKRAVLAVH DVQTLSSQLAVTVGPGERRIGPGEPELLELCNVSGALP PAGRHAAYSVGWEMAPAGAPGPGRLVAQLDTEGVGSL GPGYEGRHIAMEKVASRTYRLRLEAARPGDAGTYRCL AKAYVRGSGTRLREAASARSRPLPVHVREEGVVLEAV AWLAGGTVYRGETASLLCNISVRGGPPGLRLAASWWV ERPEDGELESSVPAQLVGGVGQDGVAEELGVRPGGGPVS VELVGPRSHRLRLHSLGPEDEGVYHCAPSAWVQHADY SWYQAGSARSGPVTVPYPMHALDTLFLVPLLVTGTGVAL VTGATVLGTITCCFMKRLRKR*
2011	A	153	1994	MGALRPTLLPPSLPLLLLLMLGMGCWAREVLVPEGPL YRVAGTAVSISCNVTGYEGPAQQNFEWFLYRPEAPDT ALGIVSTKDTQFSYAVFKSRVVAGEVQVQRLQGDVV LKIARLQAQDAGIYECHTPSTDTRYLGSYSGKVELRV LPDVLQVSAAPPGPRGRQAPTSPPRMTVHEGQELALG CLARTSTQKHTHLAVSFGRSVPEAPVGRSTLQEVVGI RSDLAVEAGAPYAERLAAGELRLGKEGTDTRYRMVVG AQAGDAGTYHCTAAEWIQDPDGSWAQIAEKRAVLAVH DVQTLSSQLAVTVGPGERRIGPGEPELLELCNVSGALP PAGRHAAYSVGWEMAPAGAPGPGRLVAQLDTEGVGSL GPGYEGRHIAMEKVASRTYRLRLEAARPGDAGTYRCL AKAYVRGSGTRLREAASARSRPLPVHVREEGVVLEAV AWLAGGTVYRGETASLLCNISVRGGPPGLRLAASWWV ERPEDGELESSVPAQLVGGVGQDGVAEELGVRPGGGPVS VELVGPRSHRLRLHSLGPEDEGVYHCAPSAWVQHADY SWYQAGSARSGPVTVPYPMHALDTLFLVPLLVTGTGVAL VTGATVLGTITCCFMKRLRKR*
2012	A	153	1994	MGALRPTLLPPSLPLLLLLMLGMGCWAREVLVPEGPL YRVAGTAVSISCNVTGYEGPAQQNFEWFLYRPEAPDT

606
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ALGIVSTKDTQFSYAVFKSRVVAGEVQVQRLQGDVAV LKIARLQAQDAGIYECHTPSTDTRYLGSYSGKVELRV LPDVLQVSAAPPGRGRQAPTSPRMTVHEGQELALG CLARTSTQKHTHLAVSFGRSVPEAPVGRSTLQEVVGI RSDLAVEAGAPYAERLAAGELRLGKEGTDYRMVVG AQAGDAGTYHCTAAEWIQDPDGSWAQIAEKRAVLAV DVQTLSSQLAVTVGPGERRIGPGEPELELLCNVSGALP PAGRHAAYSVGWEMAPAGAPGPGRLVAQLDTEGVGSL GPGYEGRHIAMEKVASRTYRLRLEAARPGDAGTYRCL AKAYVRGSGTRLREAASARSRLPVHVREEGVVLEAV AWLAGGTVYRGETASLLCNISVRGGPPGLRLAASWWV ERPEDGELESSVPAQLVGGVGQDGVAEGLVVRPGGPPVS VELVGPRSHRLRLHSLGPEDEGVYHCAPSAWVQHADY SWYQAGSARS GPVTVPYPMHALDTL FVPLLVTGTGVAL VTGATVLGTITCCFMKRLRKR*
2013	A	1273	480	YLRLWLRFDPDRHPHGVPLPTEPSTPKSPSAGPSPHL LHPGTPGHPSASPPSRPPSSSTPKRPRTAGRNPKRRQ SSPGRPT/NPGLRKKMGPPSEG\SGGGNTPOGPASGP ASLLPNPC*LCRGKPLGVLRGGGRRGASVPESWPHIP APNAG*GHAQRDPGGAGQPKD*GGRGAPGQQATEADS GPAA\GMRGPHIQLDTPLSASRGMRNARGTFGM/PS LPRGDLSPSSAGHPPASVTLPQGPFPKGTLPAGTLP PALFGDQEL
2014	A	853	1553	KKKETVSVSSREVRETSKALERPKLQE*PRGPALQSR ATSPRNTYQRPAGWPQAEPPQ*GNRLFPAQVRGRAPG PHPRA*WSQPPAEDPTGRAETQLCPAALARAQPRRQ LCGPALPGRRP/PTRTPT*SGRGFSKWLAPETITQGP APN\PFGFSVDVLFVFFKPFSLFR**KNL*KTLTLLNQ PEPQEPKGC GGVRPHYVSGLLPTLKPCSLKREGPRP ALPPS/SPSPPLCPSLRSPASL/PPVILA FRVPWR FP*PPVKIQRLSPFFNFNDN*/PSVSFSKFYFSNHPG QPPALIPSRPGLSGPPFHTLRFETAVFPTFAAGMAVS CPCLPIWPI PQPWGPGSLPQPPPLMP*KLGPRPCWP EPQMPSSGSLT/SGPNSSGLGIGPPYPGSPWGWQ*KG KAFILANRPHHPLLPGPPCRDLGLSLP/RPLLSVCGSR TLCPSPGASAVTRLLKMNS*ILPAHPRDPWSWPSS PVPETSTP*R*TLGPPTSRTCRPEV\PWALPPANWAT SFPPLTLG/VPHPLQGDYSPDPTPVSPHGPLLN
2015	A	527	871	VWSPDRPSSSDPRGQRRRPTGRVAADPGAAPPAAAAA PPPSSA*TAPGSCRRWRTSSRRPTPGSNPRPTPPRPR SRATSP/TPDSAQRLPPPPPPAGPG\PPGPEAPPVSL GQPFCR
2016	A	17	941	PLDRAVEFAVGSGRPRRISCLSCPGGGGAASGLQRAA GGTGLSWVPAGLRVCCSQRSEPEKEEQPVQNPRRKG KGGEISTWKNSSMKMKECLRIKER*TMKNSHRTRESQ K*LVFWKTRS*KTRETQKTRARELRNR*RIKKSQVR ERQKEKESQRGRESQRCREDQRQRESQREGEGQVRKE SQTWVREPESEGEPESETRAAGKRP AEDDIPRKAKRK TNKGLAQYLKQYKEAIHDMNFSNEDMIREFDNMARVE DKRRKSKQKLGAFLWMQRNLQDPFYPRGPREFRGGCR

607
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				APRRDTEIPYV
2017	A	335	120	MFLLLFCLMFDFTKVFFILLHIFCLSTCLFLGLHIC ASFHARALLEATALILLRMKIAGFQVILFPQDFVL*
2018	A	3	800	FVLDPYSGVIKSNVSFDREQQSSYTFDVKATDGGQPP RSSTAKVTINVMVDVNDNSPVVISPPSNTSFKLVPLSA IPGSVVAEVFAVDVDTGMNAELKYTIVSGNNKGLFRI DPVTGNITLEEKPAPTDVGLHRLVNVNISDLGYPKSLH TLVLVFLYVNDTAGNASYIYDLIRRTMETPLDRNIGD SSQPYQNEDYLTIMIAIIAGAMVVIVVIFVTVLVRCR HASRFKAAQRSKQGAEWMSPNQENKQNKKKRKKRKS PKSSLLN
2019	A	1	1331	GWNGSWNDNLVDTSPDKRDPLQDICRRYMEDLKKICF YRELNSKTTLKFFVHTSFHGVGHYVQLAFKVFQFKPP IPVPEQKDPDPDFSTVKCPNPEEGESVLELSRLAEK ENARVVLATDPDADRLAAAEELQENGCKVFTGNELAA LFGWWMFDCWKKKNSRNADVKNVYMLATTVSSKILKA IALKEGFHFEETLPGFKWIGSRIIDLLENGKEVLFAP EESIGFLCGTSVLDKDGVSAAVVVAEMASYLETMNT LKQQLVKVYEKYGYHISKTSYFLCYEPPTIKSIFERL RNFDSPKEYPKFCGTFAILHVRDVTGTYDSSQPNKKS VLPVSKNSQMITFTFQNGCVATLRTSGTEPKIKYYAE MCASPDQSDTALLEEELKKLIDALIENFLQPSKNGTG SGRSCLGVPNTVMTLCGAYGNRATRRNCHTLEPCG
2020	A	1	2337	TRFRGLRPAVAPWTALLALGLPGWVLAVSATAAAVVP EQHASVAGQHPLDWLLTDRGPFHRAQEYADFMERYRQ GFTTRYRIYREFARWKVNNLALERKDFFSLPLPLAPE FIRNIRLLGRRPNLQQVTENLIKKYGTHFLLSATLGG EESLTIFVDKQKLGRKTETTGGASIIIGSGNSTAVSL ETLHQLAASYFIDRESTLRLHHIQIATGAIKVTETR TGPLGCSNYDNLDSVSSVLVQSPENKVQLLGLQVLLP EYLRERFVAAALSYITCSSEGELVCKENDCWCKCSPT FPECNCPDADIQAMEDSLLQIQDSWATHNRQFEESEE FQALLKRLPDDRFLNSTAISQFWAMDTSLQHRYQQLG AGLKVLFFKTHRIILRLFNLCRCHRQPRFRLPKERS LSYWWNRIQSLLYCGESTFPGTFLEQSHSCTCPYDQS SCQGPIPCALGEGPACAHCAPDNSTRCGSCNPGYVLA QGLCRPEVAESLENFLGLETDLDLELKYLLQKQDSR IEVHSIFISNDMRLGSWFDPSWRKRMLLTLSKNKYKP GLVHVMLALSLLQICLTKNSTLEPVMAIYVNPFGGSHS ESWFMPVNEGSFPDWERTNVDAQAQCNWTITLGNRW KTFETVHVYLSRIKSLDDSSNETIYYEPLMTDPS KNLGYMKINTL\QVFGYSLPFDPD\AIRDLILQLDYP YTQGSQDSALLQLIELRDRVNQLSPPGKVRDLDFSC LRHRLKLANNEVGRIQSSLRAFNSKLNPVEYETGKL CS
2021	A	161	547	PAGIGRSTAKTPGTPGSLEMENLKSGVYPLKEASGCP GADRNLVYSFYEKGPLTFRDVAIEFSLEEWQCLDTA QQDLRKVMLENYRNLVFLAGIAVSKPDLITCLEQK EPWNMKRHAMVDQPPGR
2022	A	161	547	PAGIGRSTAKTPGTPGSLEMENLKSGVYPLKEASGCP

608
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GADRNLLVYSFYEKGPLTFRDVAIEFSLEEWDCLDTA QQDLRKVMLENYRNLVFLAGIAVSKPDLITCLEQKG EPWNMKRHAMVDQPPGR
2023	A	3	452	AVPGPGFGLSPTMVTLAELLVLLAALLATVSGY\FVS IDAHAEECFERVTSGTKMGLIFEADGGFLDIDVVI TLPDR/RKIKPRLLKKKGQ*TYRSFMDVTFKLCYNLR MSWMNPNI RNHNHLLLSIKFLITQFRSSLSYLSSC IQSE
2024	A	31	1312	ITTVMAGKRSGWSRAALLQLLLGVN LGVMPPTRARS RFVTL L YRHGDRSPVKTPKDPYQEEW PQFGQLTK EGMLQHWELGQALRQRYHGFLNTSYHRQEVYVRSTDF DRTLMSAEANLAGLFPNGMQRFNPNI SWQPIPVHTV PITEDRLLKFPLGPCPRYEQ LQNETRQTPEYQNESSR NAQFLDMVANETGLTDLTLETVWNVYDTLFC EQTHGL RLPPWASPQTMQRLSRLKDFSFRFLFGIYQQA EKARL QGGVLLAQIRKNLTLMATTSQLPKLLVYSAHDTTLVA LQMALDVYNGEQAPYASCHIFELYQEDSGNFSVEMYF RNESDKAPWPLSLPGCPHRCPLQDFLRLTEPVVPKDW QQECQLASGPADTEVIVALAVCGSILFLLIVLLLTVL FRMQAQP PGYRHVADGEDHA
2025	A	2	317	FVDSPRFRATIDEVETDVVEIEAKLDKLVKLCSGMVE AGKAYVSTSR L FVSGVRDLSQQCQGD TVISECLQRFA DSLQEVVNYHMILFDQAQRSVRQQQLQSFVKE
2026	A	1788	3	RTRGRFPKRTP/LFQISSAVQKEQPLPTAEITRLAVW AAVQAVERKLEAQAMRLLTLEGRTGTNEKKIADCEKT AVEFANHLESKWVVLGTL LQ EYGLLQRRLENMENLLK NRNFWILRLPPGSNGEVPKVPVTFDDVAVHFSEQEWG NLSEWQKELYKNVMRGNYESLVSM DY AISKPDLM SQM ERGERPTMQEQEDSEGETPTDPSAAHDGIVIKIEVQ TNDEGESLETPEPLMGQVEEHGFQDSELGDPCGEQP DLDMQEPENTLEEST/DRLQRVQRTEADAGAAEELHG /VGS/WIKTEEQDEEEEEEEDELPQHLQSLGQLSGR YEASMYQTPLPGEMSPEGEESPPPLQLGNPAVKRLAP SVHGER/PPERPRGLEPAAAEPARRAALHMHGVRQE LPP/GRSTSSSTSATT SRRGPTSAPNARSASGTSNSS RCTSASTACAEAA SHPN/CGPTFNPKHALKPRPKSPS SGSGGGGPKPKCPECDSSFSHKSSLTKHQITHTGER PYTCPECKKSFR L H I S L V I H Q R V H A G K H E V S F I C S L C GKSFSRPSHLLRHQRTHTGERPFKCPECEKSFSEKSK LTNHCRVHSRERP
2027	A	2193	442	ELNCNIRAPPKQMFWCFRPRSKERAVVVAWERRLMVV GDAPESI QFVLDEDSYLVPELDGVRIFSRSTHEFLHE VPAASEE I F K I A S M A P G A L L L E A Q K E Y E K E S Q K A D E Y LREIQELGQLTQAVQQCIEAAGHEHQPD MQSLLRAA SFGKCF LDRFPD S F V H M C Q D L R V L N A V R D Y H I G I P L TYSQYKQLTIQVLLDRLVLRRLYPLAIQICEYLR LPE VQGVSRILAHWACYKVQKDVSD EDVARAINQKL GDT PGVSYSDIAARAYGCGRTELAIKLLEYEPRS GEQVPL LLKMKRSKLALSKAIESGDTDLVFTVLLHLKNE LNRG DFFMTLRNQPMALSLYRQFCKHQELET LKDLYNQDDN

609
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				HQELGSFHIRASYAAEERIEGRVAALQTAADAFYKAK NEFAAKATEDQMRLRLQRRLEDELGGQFLDLSLHDT VTTLILGGHNKRAEQLARDFRI PDKRLWWLKL TALAD LEDWEELEKFSKSKSPIGYLPFVEICMKQH NKYEAK KYASRVGP EQVKALLLVGDVAQAADVAIEHRNEAEL SLVLSHCTGATDGATADKIQRARAQAQKK
2028	A	110	277	MLLALPLAAPSCPMLCTCYSSPPTVSCQANNFSSVPL SLPPSTQRLFLQNNLIRTL
2029	A	1	359	ISGESIYWSQKPTPSSNASPWSEPAAVDVELTAYALL AQLTKPSLTQKEIAKATSIVAWLAKQRNAYGGFSSTQ DTVVALQALAKYATTAYVPSEEINLVVKSTENFQRTF NIQAVNRM
2030	A	16	255	ARPSCPCSWSFSCCGVSPGA/LVTEAAIFYETQPSLW AESESLKPLAKLMTYFKNSTYLIRLFMIYRCKPVKS KKKKRN
2031	A	2	414	GKTHTATVVELNPWVEYEFVVASNKIGGGEPSPSE KVRTEEAVPEVPPSEVNGGGGSRSELVITWDPVPEEL QNGEGFGYVVAFRPLGVTTWIQT VVTSPDTPRYVFRN ESIVPYSPIYEVKVG VYNNKGE GPFSP
2032	A	3	438	SNLHHLILN NNQLT LISSTAFDDVFAL EELDLSYNNL ETIPWDAVEK MVS LHTLSLDHNMIDNIPKGTFSHLHK MTRLDVTSNKLQKLPPDPLFQRAQVLATSGIISPSTF ALSFGGNPLHCNCELLWLRRLSREDDLET CASPP
2033	A	3	438	SNLHHLILN NNQLT LISSTAFDDVFAL EELDLSYNNL ETIPWDAVEK MVS LHTLSLDHNMIDNIPKGTFSHLHK MTRLDVTSNKLQKLPPDPLFQRAQVLATSGIISPSTF ALSFGGNPLHCNCELLWLRRLSREDDLET CASPP
2034	A	166	4280	ASDQSGSQPGDHSAGQANQLKLEDMKSPRRTTLCLMF IVIYSSKAALNWN YESTIHPLSLHEHEPAGEEALRQK RAVATKSPTAE EYTVNIEISFENASFLDPIKAYLNSL SFPIHGNNTDQITDILSINVTTVCRPAGNEIWCSCET GYGWPRERCLHNLICQERDVFLPGHHCSCLKELPPNG PFCLLQEDVTLNMRVRLNVGFQEDLMNTSSALYRSYK TDLETAFRKGYGILPGFKGVTVTGFKSGSVVVTYEVK TTPPSLELIHKANEQVVQSLNQTYKMDYNSFQAVTIN ESNFFVTPEIIFEGDTVSLVCEKEVLSSNVSWRYEEQ QLEIQNSSRFSIYTALFNNMTSVSKLTIHNITPGDAG EYVCKLILDI FEYECKKKIDVMP IQILANEEMKVMCD NNPVSLNCCSQGNVNWSKVEWKQEGKINIPGTPETDI DSSCSRYTLKADGTQCPSGSSGTTVIYTCEFI SAYGA RGSANIKVTFISVANLTITPDPI SVSEGQNF SIKCIS DVSNYDEVYWN TSAGIKIYQRFYTTRYLDGAESVLT VKTSTREWNGTYHCIFRYKNSYSIATKDVI VHPLPLK LNIMVDPLEATVSCSGSHHIKCCIEEDGDYKVT FHMG SSSLPAAKEVNKKQVCYKHNFNASSVSWCSKTVDVCC HFTNAANNSVWSPSMKLNLPGENITCQDPVIGVGEP GKVIQKLCRFSNPSSPEE/SPLGGTITYKCVGSQWG \EKRND C I SAPINSL LQMAKALIKSPSQDEMLPTYLK DLSISIDKAEHEISSSPGSLGAIINILDLLSTVPTQV NSEMMTHVLSTVNVLGKPV LNTWKVLQQQWTNQSSQ

610
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				LLHSVERFSQALQSGDSPPLSFSQTNVQMSSTVIKSS HPETYQQRFFVFPYFDLWGNVVIDKSYLENLQSDSSIV TMAFPTLQAILAQDIQENNFASLVMTTTTVSHNTTMP FRISMTFKNNSPSGGETKCVFNFRLANNTGGWDSSG CYVEEGDGDNVTICDHLTSFSILMSPDSPDPSSLLG ILLDIISYVGVGFSILSLAACLVVEAVVWKSVTKNRT SYMRHTCIVNIAASLL\ VANTWFIGVAAIQDNRYILC KTACVAATFFIHFYLSVFFWMLTLGLMLFYRLVFIL HETSRSTQKAI AFCLGYGCPLAISVITLGATQPREVY TRKNVCWLNWEDTKALLAFAPALIIVVNITITIVV ITKILRPSIGDKPKCKQEKSSLFQISKSIGVLTPLLGL TWGFGLTTVPFGTNLVFHIIIFAILNVFQGLFILLFGC LWDLKVQEALLNKFSLSRWSSQHSKSTSLGSSTPVFS MSSPISRRFNNLFGKTGTYNVSTPEATSSSLENSSSA SSLLN
2035	A	1	366	AFRSDSRLAEHQRVHTGERPYTCNECGKVFSTKAYLA CHQKLHTGEKLYECECDKVYIRKSHLERHRIHTGE KPHKCGDCGKAFNSPSHLIRHQRIHTGQKSYKCHQCG KVFSLRSLLA
2036	A	2	236	ISGOEGLQAVLASDYSFAQFRYLQRLLLVHGRWSYFR MCKFLCYFFYKNFAFTLVHFWFGFFCGFSAQTVYDQW FITL
2037	A	706	951	MRCGWGPLGCLGTGAPAGWMVLGSPRSQQLQARWSRA SLSAFGWEIRLRPEGPKAPRQLLLVALESETLGVHGG ATPLHCL*
2038	A	1242	433	PGSPDVNRAVVRPPPPPPPPPPAPQPTMSRRKQGKPO HLSKREFSPEPLEAILTDDEPDHGPLGAPEGDHDLLT CGQCQMNFPGLDILIFIEHKRKQCNGSLCLEKAVDKP PSPSPIEMKKASNPVEVGIQVTPEDDDCLSTSSRGIC PKQEHIAKLLHWRGLSSPRSAHGALIPTPGMSAEYA PQGICKDEPSSYTCTTCKQPFTSAWFLLOHAQNTGL RIYLESEHGSPLTPRVLHTPPFGVVPRELKMGCSFRM EAREPLSSEKI
2039	A	2009	1889	MHSAMLGTRVNLVSDFWRVMMRVCWLVRQDSRHQRI RLPHLEAVVIGRGPETKITDKKCSRQVQLKAECNKG YVKVKQVGVNPTSIDSVVIGKDQEVKLQPGQVLHVMN ELYPYIVEFEEEEAKNPGLETHRKRKRSGNSDSIERDA AQEAEAGTGLEPGSNSGQCSVPLKKGKDAPIKKESLG HWSQGLKISMQDPKMQVYKDEQVVVIKDKYPKARYHW LVLPWTSISSLKAVARGTP*TP*AYAHCGGKGDRCFC W\SSKLRFRLLGYHAIPSMHVHLHVISQDFDSPCLKN KKHWNSFNTEYFLESQAVIEMVQEAGRVTVRDGMPEL LKLPLRCHECQQLLPSIPQLKEHLRKHWTQ*FFFFT VLSKFILREKESSTQLFHSPTTFPCIRTYAVIVS
2040	A	2009	1889	MHSAMLGTRVNLVSDFWRVMMRVCWLVRQDSRHQRI RLPHLEAVVIGRGPETKITDKKCSRQVQLKAECNKG YVKVKQVGVNPTSIDSVVIGKDQEVKLQPGQVLHVMN ELYPYIVEFEEEEAKNPGLETHRKRKRSGNSDSIERDA AQEAEAGTGLEPGSNSGQCSVPLKKGKDAPIKKESLG HWSQGLKISMQDPKMQVYKDEQVVVIKDKYPKARYHW

611
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LVLPWTSISSLKAVARGTP*TP*AYAHC GGK GDCRFC W\SSKLRFR LGYHAI PSM SHVHLHVI SQDFDSPCLKN KKHWN SFNTEYFLESQAVIEMVQEAGRVTVRDGMPEL LKLPLRCH ECQQLLPSI PQLKEHLRKHW TQ*FFFFTV LSKFILREKESSGSTQLFHSPTTFPCIRTYAVIVS
2041	A	2009	1889	MHSAMLGTRVNL SVSDFWRVMMRVCWLVRQDSRHQRI RLPHLEAVVIGRG PETKITDKKCSRQQVQLKAECNKG YVKVKQVGVNPTSIDS VVIGKDQEVKLQPGQVLHMVN ELYPYIVEFE EEEAKNPGLETHRKRKRSGNSDSIERDA AQEAEAGTGLEPGS NSGQCSVPLKKGKDAPIK KESLG HWSQGLKISMQDPKM QVYKDEQVVVIKDKYPKARYHW LVLPWTSISSLKAVARGTP*TP*AYAHC GGK GDCRFC W\SSKLRFR LGYHAI PSM SHVHLHVI SQDFDSPCLKN KKHWN SFNTEYFLESQAVIEMVQEAGRVTVRDGMPEL LKLPLRCH ECQQLLPSI PQLKEHLRKHW TQ*FFFFTV LSKFILREKESSGSTQLFHSPTTFPCIRTYAVIVS
2042	A	1464	775	KMTTAARPTFEPARGGRGKGEGDLSQLSKQYSSRDLP SHTKIKYRQT TQDAPEEVRNRDFRRELEERERAAARE KNRDRPTREHTTSSSVSKKPRLDQI PAANLDADDPLT DEEDED FEEESDDDDTAALLAELEKIKKERAEEQARK EQEQKAE EERIRMENILSGNPLLNL TGPSQPQANFKV KRRWDDDV FKNCAKGVDDQKKDKRFVNDTLRSEFHK KFMEKYIK
2043	A	2	860	ATTRIRLSGGRSQHEGRVEVQIGGPGPLRWGLICGDD WGTLEAMVACRQLGLGYANHGLQETWYWD SGNITEVV MSGVRCTGT ELSLDQCAHHGTHITCKRTGTRFTAGVI CSETASDLLLHSA LVQETAYIEDRPLHMLYCAA EENC LASSARSANWPYGHRRLLRFSSQIHN LGRADFRPKAG RHSWVWHECHGHYHSMDFFTHYDILTPNGTKVAEGHK ASF CLEDETECQEDVSKRYECANFGEQGITVGCWDLYR HDIDCQWIDITDVKPGNYILHGVINPT
2044	A	973	266	ARGSLCAPASPLYPNQLRNVALAQALTPYVFLSDID FLPAYSLYDYL RASIEQLGLGSRRKAALVVP AFETLR YRFSFPHSKVELLALLDAGTLYTFRYHEWPRGHAPTD YARWREAQAPYRVQWAANYEPYVVVPRDCPRYDPRFV GFGWNKVAHIVELDAQEYELLVLPEAFTIHLPHAPSL DISRFRSSPTYRDCLQALKDEFHQDLSRHHGAAALKY LPALQQPQSPARG
2045	A	1668	218	AVVRAQGSRGFSGAGWRPRQAAAMNFSEVFKLSSLLC KFSPDGKYLASC VQYRLVVRDVNTLQILQLYTCLDQI QHIEWSADSLFILCAMYKRGLVQVWSLEQPEWHCKID EGSAGLVASCWSPDGRHILNTTEFHLRITVWSLCTKS VSYIKYPKAC LQGITFTRDGRYMALAERRDCKDYVSI FVCSDWQLLRHFD TDTQDLTGIEWAPNGCVLAVWDTC LEYKILLYSLDGRLLSTYSAYEWSLGIKSVAWSPSSQ FLAVGSYDGKVRILNHVTWKMIT EFGHPAAINDPKIV VYKEAEKSPQLGLGCLSFPPPRAGAGPLPSESKEYEI ASVPVSLQTLKPVTD RANPKIGIGMLAFSPDSYFLAT RNDNIPNAVWWDIQKLRLFAVLEQLSPVRAFQWDPQ QPRLAICTGGSRLYLWSPAGCMSVQVPGE GDFAVLSL

612
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				CWHLSGDSMALLSKDHFCLCFLETEAVVGTACRQLGGHT
2046	A	231	1289	SPTVSFLFFNMETNPSVGTTSALISILLARSSRERQLS SEGRFSWRL*DASSGERS*RRSESSSWLSS*ERESSV SFKHPFKRLFK*SSVSLLSWSSLSPFSSGAIHTSGSS MPKSDI*LFPQSTFSEPSESACACGDFPSLSVRSGCC SSFNSLFSSWSVGNASEASRSGKRSSF*ACEYLPSE INAGGIRSQPGEINGSVFDLLERNTLGSSAMPSILAT SWQASV*ASCKRLSSSQASSEESGPDGLPAVSEDWVW SANVASALQSSSSMWSFPAVTERLGESVC\SPSDDSR DCSPGAPLYVGFLYLTLCDKFKYSLKMKKNKLLKIQN NTLYRKEKKGHMNMNMCNTAIF
2047	B	26	175	NCGSGDILLKIVKVEHEEMPEAKNVIADVLEEFMKEAL DQSF
2048	A	1	1386	RDFVAASSRRRRADFPRMTELQRVAHEPVAPPEDKE SESEAKVDGETASDESRAESAPLFVSADDTPEVLNR ALSNLSSRWKNWWVRGILTAMIAFFFFIIYLGPMVL MIIVMCVQIKCFHEIITIGYNYHSYDLFWFRTLSWY FLLSVNYFFYGETVTDYFFTLVQREEPLRILSKYHRL ISFTLYLIGFCMFVLSLVKKHYRLQFYMFGWTHVTLL IVVTQSHLVIHNLFEGLMIWFIVPISCVICNDIMAYMF GFFFGRTPLIKLSPKKTWEGFIGGFFATVVFGLLLSY VMSGYRCFVCPVEYNNDTNSFTVDCEPSDLFRLQEYN IPGVIQSVIGWKTVMYPFQIHSIALSTFASLIGPFG GFFASGFKRAFKIKDFANTIPGHGGIMDRFDCQYLMA TFVNVYIASFIRGNPNPSKLIQQFLTLRPDQQLHIFNT LRSHLIDKGMLTSTTEDE
2049	A	2	427	HSWVSRSCAFEPAAWEEGATSQTVATCGGEAVCVIDCQ TGIVLHKYKAPGEEFFSVAWTALMVVTQAGHKRWSV LAAAGLRGLVRLHLVRAGFCCGVIRAHKKAIALTLCS PAHETHLFTASYDKRIILWDIGVNPQDYEFQ
2050	A	1	892	RTRGRTRGRGTRGGGGGGTGAGGRGEGSQVPGLSAA DQDR*GRGCCSPGGRDRAGGGGGIGQGGDAERRRGEQ GEGWGRTPGQKPGRGEPALWKGRV*GPRVVRGGPEAA GAAAQRPPGPVFPFAGGAEPALPALQPIPAQDLRGA AQKEGPGGR*GG*PGRRGRGPRERASVPAPSGHAGGA EEAAGRPAVPPGAGPVEAAVPGEAHQGGEGVATLP GTQEAGGDAGHGQLSDEGRAPGCSARGGADPGVGG*K GEGDERRAAGEHSAEAPGAF*NQDEDPGGPDPGSAS Y
2051	A	2	1086	FVLCAGACWPLRDRDT/SPPAHLCPVTPWSLHVPIS LQCPRLCSPPTHRLTPPAGCQRPPAGPLSVAPASL SPSAPALLEA/TSPPWTAGATWSPGRSPATQCWPPSW CQTPFPHPETGQLCLVRS LH*PHLSSLGQAGAAG*GG PLAPPFPFLVPFP\P*QVQHPRSPA*GAGPEPAVNI PQPL/PVPPWD*PLTSPPNSTGAPSWPRAGSVSPSP/ VLEPRPEQLSGRQGCSSVSSWGAPGGATDRQAAQGP HPSPGRCCPRRTVLGNEPPAGFGLRSLWPRSPPEVG ARLPNGAFGFSVRCLLCFPPWRAEPPHIRIGRATPPG PGP/VPSQPSPRGSMVPRPGAARGQLDGHVQGSRL

613
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2052	A	3	1385	KYESAQPGGTQPEPGLGARMATHKALVMCLGLPLFLF PGAWAQGHVPPGCSQGLNPLYYNLCDSGAWGIVLEA VAGAGIVTTFVLTIILVASLPFVQDTKKRSLLGTVF FLLGTLGLFCLVFACVEKPDFSTCASRRFLFGVLF CFSCLAHVFAFNFLARKNHGPRGWVIFTVALLTLV EVIINTEWLIITLVRGSGEGGPQGNSSAGWAVASPCA IANMDFVMALIYVMLLLGAF LGAWPALCGRYKRWRK HGVFVLLTTATSVAIWVWVIMYTYGNKQHNSPTWDD PTLAIALAANAWAFVLFYVIPEVSQVTKSSPEQSYQG DMYPTRGVGYETILKEQKQSMFVENKAFSMDEPVAA KRPVSPYSGYNGQLLTSVYQPTMALMHKVPSEGAYD IILPRATANSQVMGSANSTLRAEDMYSAQSHQAATPP KDGKNSQVFRNPYVWD
2053	A	2	555	MASPAASSVRPPRPKKEPQTLVIPKNAAEEQKLEKLER LMKNPDKAVPIPEKMSEWAPRPPPEFVRDVMGSSAGA GSGEFHVYRHLRRREYQRQDYMDAMAEKQKLDAEFQK RLEKNKIAAEEQTAKRRKKRQKLKEKKLLAKMKLEQ KKQEGPGQPKEQGSSSSAEASGTEEEEEEVPSFTMGR
2054	A	1008	534	HEKMAAAWGSSSLTAATQRAVTPWPRGRLLTASLGPOA RREASSSSPEAGEGQIRLTDSCVQRLLLEITEGSEFLR LQVEGGGCSGFQYKFSLDTVINPDDRVEQGGARVVV DSDSLAFVKGAQVDFSQELIRSSFQVLNNPQAQGCSC CGSSFSIKL
2055	A	1492	528	THVVMTGMCYAPHQVLSYINGVTTSKPGVSLVYSMP RNLRLRLEGLQEKDSGPYSCSVNVQDKQKSRGHSIK TLELNVLPVPPAPPSCLQGVPHVGANVTLSQSPRSK PAVQYQWDRQLPSFQTFAPALDVIRGSLSLTNLSS MAGVYVCKAHNEVGTAQCNVTLEVSTGPGAAVVAGAV VGTLVGLGLLAGLVLLYHRRGKALEEPANDIKEDAIA PRTLWPWKSSDTISKNGTLSSVTSARALRPPHGP GALTPTPSLSSQALPSPRLPTTDGAHPQPISPIPGGV SSSGLSRMGAVPVMVPAQSQAQAGSLV
2056	A	820	319	VVEFPVLTKAATSGILSALGNFLAQMIKKRKKENSR SLDVGGPLRYAVYGGFFFTGPLSHFFYFFMEHWIPPEV PLAGLRRLLLDRLVFAPAFMLLFFLIMNFLEGKDASA FAAKMRGGFWPALRMNWRVWTPLOFININYVPLKFRV LFANLAALFWYAYLASLGK
2057	A	520	330	HGCVLSLLPKPQQGFREPVLHSTC/PNPTPPVPP*S DRYLSNPTQPVPP*SDRYLSNPTPPVSP*SDRYLSNP TPPVPP*SDRYLSNRTPPVSP*SDRYLSNPTPPVSP
2058	A	2	479	DTGQKGLPGPPGPPGYGSQGIKGEQGPQGFPGPKGTM GHGLPGQKGEHGERGDVGKKGDKGEIGEPGSPGKQGL QGPKGDLGLTKEEIIKLITEICGCGPKCKETPLELVF VIDSSESVGPENFQIIKNFVKTMADRVALDLATARIG IINYSHKVEKV
2059	A	503	1051	VFLYPFLKWWRDP*RRELPTFWFLELAIFTLIEEV LFYYSHRLLHHPTFYKKIHKKHHEWTAPIGVISLYAH PIEHAVSNMLPVIVGPLVMGSHLSSITMWFSLALIIT TISHCGYHLPFLPSPEFHDYHHLKFNQCYGVLGVL LHGTDTMFKQTKAYERHVLLLGFTPLSESIIPDSPK

614
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2060	A	1	716	ERVGNVCSLEISNIQKGEYMCCHAVNIIGEAKSFA NVDIMPQ/RRKSGGTTTSR/IFVDPNMDSREGEDKEL KIDLEVFEPPRFIMPICDFKIPENSDAVFKCSVIGI PTPEVKWYKEYMCIEPDNIKYVISEEKGSHTLKIRNV CLSDSATYRCRAVNCVGEAICRGFLTMDSEIFA VIAKKSQVTLSSLMEEVLKSNYTDSSFFEFQVGE GPPRFIKGISDCYAPIGTAAAYFQCL
2061	A	47	538	RVRLRPVFCVMTSQEKTEEYPFADIFDEDETERN FLLSKPVCVVFVGKPGVGKTTLARYITQAWKCI RVEALPILEEQIAAETESGVMLQSMILSGQSI PDELVIKLMLEKLNSEVCHFGYIITEIPSLSQD AMTTLQQIELIKNL\NLKPDVLIINIKGVLD F
2062	A	1196	230	RARSGLQGA VPLGPTGRSRHSLQTKLPSSPF SERPLVFQTPGALVSTPHGRYPPLCPPKAAFQ KVIHGKAVPSNPS/VVPTAIVNPVRSTAGP GTLGQGSRLRKRSSMRKNGSLQRPLQSGI PTLVVGSLLRRSPT/MGPSASAVPILPATG DPLLPLSRGGGDGVQA/SPSRGSPSRASAG AVRPGSTPRPAPSLWKTCKKSPSRVSLCQNR PHLPHHPSW*NQKTQEMASKSKSKP*DFRIT ALLPPNITPPIPPP/AKPEQPATLKASQPEA ASLGPEMTVLFARHSGCHSGQQTDLRRKSA LGKATTLVSTASGTQTVFPSK
2063	A	1196	230	RARSGLQGA VPLGPTGRSRHSLQTKLPSSPF SERPLVFQTPGALVSTPHGRYPPLCPPKAAFQ KVIHGKAVPSNPS/VVPTAIVNPVRSTAGP GTLGQGSRLRKRSSMRKNGSLQRPLQSGI PTLVVGSLLRRSPT/MGPSASAVPILPATG DPLLPLSRGGGDGVQA/SPSRGSPSRASAG AVRPGSTPRPAPSLWKTCKKSPSRVSLCQNR PHLPHHPSW*NQKTQEMASKSKSKP*DFRIT ALLPPNITPPIPPP/AKPEQPATLKASQPEA ASLGPEMTVLFARHSGCHSGQQTDLRRKSA LGKATTLVSTASGTQTVFPSK
2064	A	1554	1358	EFVMRHKGAKHLRSAHDLTWFQHYSDVIGFL LTCVATAIFLFTKCFLESCQKFNKTRKIEKRE
2065	A	793	279	HEGASLGVRGGGMADTVLFEFLHTEMVAEL WAHDPDPGPGGQKMSLSVLEGMGFRVGQAL GERLPRETLAFREELDVLKFLCKDLWVAVFQ KQMDSLRTNHQGTIVLQDNSFPLLLPMASGL QYLEEAPKFLAFTCGLLRGALYTLGI ESVVTASVAALPVCKFQVVIKPS
2066	A	729	487	IIFIYLFIFLRWSL/GSVAQAEVQWPHLNSLQ APPPGFAPFSCRLRPSSWDYRHLPPCPANFL YFWWRRGFTMLARMVLIS*PRDPPASASQGA GIAGMSHCARP*MNYFYLFYFFEMESRSVAQ AEVQWPHLNSLQAPPPGFAPFSCRLRPSSWD YRHLPPCPANFLYFWWRRGFTMLARMVLIS
2067	A	1	692	PGGNRSSSSSSCRRCICTFCTCRSRRRRRSHQ PRRSSWGPLQAEVRLEFPSEKRRSGSTRGGRG GSTGVASVGSSTWGGTPGLGQTGTWQG/HTG QRGPQLPPHP\RNSFSSRHRGSSG\RLSQA\LP EPRGLESKGTGSARGVAAGRHQEGEAATGGG PRDIAQQGGCRGSACGRRSHEALRPRVWCGE GPQWTW\CAVCPHRSAPGAGLAD\RQHPGESRA WGETRLGEAGGAE

615
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
2068	A	114	1031	MPLLTLYL LLLFWLSGYSIATQITGPTTVNGLERGS LT VQCVYRSGWETYLKWWCRGAIWRDCKILVKTS GSEQE VKRDRVSIKDNQKNRTFTVTMEDLMKTDADTYWCGIE KTGNDLGVTVQVTIDPASTPAPTTPSTTFTAPVTQE ETSSSPTLTGHHLNDRHKLLKLSVLLPLIFTILL LLLL VAASLLAWRMMKYQQKAAGMSPEQVLQPLEGDL CYAD LTLQLAGTSPQKATTKLSSAQVDQVEVEYVTMASLPK EDISYASLT LGAEDQEPTYCNMGHLSSHLPGRGP EEP TEYSTISRP*
2069	A	114	1031	MPLLTLYL LLLFWLSGYSIATQITGPTTVNGLERGS LT VQCVYRSGWETYLKWWCRGAIWRDCKILVKTS GSEQE VKRDRVSIKDNQKNRTFTVTMEDLMKTDADTYWCGIE KTGNDLGVTVQVTIDPASTPAPTTPSTTFTAPVTQE ETSSSPTLTGHHLNDRHKLLKLSVLLPLIFTILL LLLL VAASLLAWRMMKYQQKAAGMSPEQVLQPLEGDL CYAD LTLQLAGTSPQKATTKLSSAQVDQVEVEYVTMASLPK EDISYASLT LGAEDQEPTYCNMGHLSSHLPGRGP EEP TEYSTISRP*
2070	A	114	1031	MPLLTLYL LLLFWLSGYSIATQITGPTTVNGLERGS LT VQCVYRSGWETYLKWWCRGAIWRDCKILVKTS GSEQE VKRDRVSIKDNQKNRTFTVTMEDLMKTDADTYWCGIE KTGNDLGVTVQVTIDPASTPAPTTPSTTFTAPVTQE ETSSSPTLTGHHLNDRHKLLKLSVLLPLIFTILL LLLL VAASLLAWRMMKYQQKAAGMSPEQVLQPLEGDL CYAD LTLQLAGTSPQKATTKLSSAQVDQVEVEYVTMASLPK EDISYASLT LGAEDQEPTYCNMGHLSSHLPGRGP EEP TEYSTISRP*
2071	A	51	1464	ALPGEFFFRFHPAHKHCHLLPPSLFTNVT TQSEISSF LSFLHFQQVPLRQKPRRKTQGFLTMSRRRISCKDLGH ADCQGWLYKKKEKGSFSLSNKWKKFWVILKGSSLYWYS NQMAEKADGFVNLPDFTVERASECKKKHAFKISHPQI KTFYFAAENVQEMNVWLNKLGS AVIHQESTTKDEECY SESEQEDPEIAAETPPPPHASQTQSLTAQQASSSSPS LSGTSYSFSSLENTVKTPSSFPSSLSKERQSLPDTVN SLSAAEDEGQPITFAVQVHSPVPSEAGIHKALENSFV TSESGFLNSLSSDDTSSLSSNHDH LTVDPDKPAGSKIM DKEETKVSEDDMEKLYKSLEQASLSPLGDRRPSTKK ELRKS FVKRCKNPSINEKLHKIRTLNSTL KCKEHDLA MINQLLDDPKLTARKYREWKVMNTLLIQDIYQQQRAS PAPDDTDDTPQELKKSPSSPSVENSI
2072	A	87	477	IKSKLNQQVEVQESEWRLTEAKGPTMGKESGWD SGRA AVAAVGGVVAVGTVLVALS AMGFTSVGIAASSIAAK MMSTAAIANGGGVAAGSLVAILQSVGAAGLSVT SKVI GGFAGTALGAWLGSPSS
2073	A	87	477	IKSKLNQQVEVQESEWRLTEAKGPTMGKESGWD SGRA AVAAVGGVVAVGTVLVALS AMGFTSVGIAASSIAAK MMSTAAIANGGGVAAGSLVAILQSVGAAGLSVT SKVI GGFAGTALGAWLGSPSS
2074	A	112	483	AGVGALRMVQRLTYRRRLSYNTASNKTRL SRTPGNRI VYLYTKKVGKAPKSACGVC PGRLRGVRAVRPKVLMRL

616
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SKTKKHVSRA YGGSMCAKCVRDRIKRAFLIEEQKIVV KVLKAQAQSQKAK
2075	A	2	446	FQNMTCELHLTCSVEDADDNV SFRWEALGNTLSSQPN LTVSWDPRI SSEQDYTCIAENAVSNLSFSVSAQKLCE DVKI QYTDTKMILFMVSGICIVFGFIILLLLVLRKRR DSLSTQRTQGPAESARNLEYVSVSPTNNTVYASVT
2076	A	1208	249	VGWSVHRYVLLHHVMGGLEGMQGAWGYVQGGMGALSD AIASSATTHGASIFTEKTVAKVQVNSEGC VQGVVLED GTEVRSKMVLSNTSPQITFLKLTPQEWLPEEFLEIRIS QLDTRSPVT KIN/V*EAHHIAALSPLTHLSEKPPGWG Q/HELSHHLH/CPDLQPVSPCSLVRSGRRQAAQ/PSW RPPMLPGASRCPI TNAPST*TVKTPSSFIRPLKMPWM ACLPTVFD CIEVYAPGFKDSVVG RDILTPDLERIFG LPGGNIFHCAMSLDQLYFARPVPLHSGYRCPLQGLYL CGSGAHPGGVMGAAGR NAAHVAFRDLKSM
2077	A	38	376	MALGVPI SVYLLFNAMTALTEEA AVTVTPPITAQQGN WTVNKTEADNIEGP IALKFSHLCLEDHNSYCINGACA FHHELEKAI CRCFTGYTGERCLKLKSPYNVCSGERRP L*
2078	A	38	376	MALGVPI SVYLLFNAMTALTEEA AVTVTPPITAQQGN WTVNKTEADNIEGP IALKFSHLCLEDHNSYCINGACA FHHELEKAI CRCFTGYTGERCLKLKSPYNVCSGERRP L*
2079	A	38	376	MALGVPI SVYLLFNAMTALTEEA AVTVTPPITAQQGN WTVNKTEADNIEGP IALKFSHLCLEDHNSYCINGACA FHHELEKAI CRCFTGYTGERCLKLKSPYNVCSGERRP L*
2080	A	1	675	MAPPLRPLARLRPPGM LLLRALLLLLLLSPLPGLREGI GELITPIGTSLPDLDPARRRWE GGI GRVGVSEVADLCP GKEGGKVPEAEKEGVWCFS ELSFVKEPQDVTVTRKDP VVLDCQAHGEVPIKVTWLKNGAKMSENKRIEVL SNGS LYISEVEGRRGEQSDEGFYQCLAMNK\F*AILNQKAH LALSRI GST*RRRPDRP*EDEAFVMTTHCFQDLLTSL IES
2081	B	1	3147	MAKISASRAEKVLEHPGEREKGREMASPWNHSLALA AVVVIISMVLLGRSIQASRKEKMQPPEKETPEVLHLD EAKDHNSLNNLRETL LSEKPNLAQVELELKERDVLSV FLPDVPETESYISVNMALPPFFGQGRPGPPPPQPPP LALFGCPPPLPSPAFPPLPQRP GPFP GASAPFLQP PLALQPRASAQASRG GGGAGAFYPVPPPLPPPPQC RPFPGTDAGERPRPPPGPGPPWSPRWPEAPPPADV LGDAALQRLRDRQWLEAVFGT PRRAGCPVPQORTHAGP SLGEVRARLLRALRLVRRLRGLS QALREAEADGA AV LLYSQTAPLRAELAERLQPLTQAAYVGEARRRLERV RRRLRLRERAREREAE REAEARAVEREQEIDRWRVK CVQEEVEKKRFFCEILTDELVLWEPSGRPQPQQLQIL TAMSTSTFYDKELKTARENKEEELIDKLEVVTMPSPS PKGLPVKQYAVQSQLPVYEWPDVGSGEYDVGVVASFG RLNEALILKFYPY SALGGSGSPAPLTRLASPAAPQDG QVDLEGRALRPAARAGFSKHRGHGDALDGHAGLRPEL

617
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				HAPLTVVADGLFSKFRKSLVSNKVSVSSHVFVGLMKH DFSLERTALFWVEAAGQGSPSPYQCGDPTASAPPAWL LLVSPEHGLAPAPTTTIRDPEAGHQERPEEEGEDEAEA SSGSEEEEPAPSSLQPGSPASPGPGRRRLCSLDVLRGVR LELAGARRRLSEGKLVSRPRALLHGLRGHRALSCLPS PAQSPRSASPPGPAPQHPAASPAPPRPSTAGAIPLLR SHKPTVAIYITTKRLPYFPFIVNFLFLIAQLPKLQYNK NVALTVKFLTKRFISEYDPNLGMVCRKPTDPVDWPPL VLGLLTLMKQFHSRYTEQFLALIGQFICSTVEQCTRQ VTKAEGVALAGRFGLFFEVSACLDFEHVQHVFEAV REARRELEKSPLTPPLFISEERALPHQAPLTARHGLA SCTFNTLSTINLKEMPTVAQAKLVTVKSSRAQSKRKA PTLTLLKGFKIF
2082	A	85	839	RSGSLMAAAAATKILLCLPLLLLLSGWSRAGRADPHS LCYDITVIPKFRPGPRWCAVQGVDEKTFLLHYDCGNK TVTPVSPLGKKLNVTTAWKAQNPVLRVVDILTQLR DIQLENYTPKEPLTLQARMSCEQKAEGHSSGSWQFSF DGQIFLLFDSEKRMWTTVHPGARKMKEKWENDKVVAM SFHYFSMGDCIGWLEDFLMGMDSTLEPSAGAPLAMSS GTTQLRATATTLILCCLLILLPCFILPGI
2083	A	1	1742	VSAVEFVLHGKDFQVDCKASGSPVP*ISWSLLDGTMI NNAMQADDSGHRTRRYTLFNNGTLYFNKVGVAEEGDY TCYAQNTLGKDEMKVHLTVITAAPRIRQSNKTNKRIK AGDTAVLDCEVTGDPKPKIFWLLPSNDMISFSIDRYT FHANGSLTINKVKLLDSGEYVCVARNPSGDDTKMYKL DVVSKPPLINGLYTNRTVIKATAVRHSHKHFDCAEAG TPSPEVMWIMPDNIFLTAPYYGSRITVHKNGTLEIRN VRLSDSADFICVARNEGGEVSVLVVQLEVLEMLRRPTF RNPFFNEKIVAQLGKSTALNCSVDGNPPPEIIWILPNG TRFSNGPQSYQYLIASNGSFIISKTTREDAGKYRCAA RNKVGYYIEKLVILEIGQKPVILTYAPGTVKGISGESL SLHCVSDGIPKPNIKWTMPSGYVVDPRQINGKYILHD NGTLVIKEATAYDRGNYICKAQNSVGHTLITVPMIV AYPPRITNRPPRSIVTRTGAAFQLHCVALGVPKPEIT WEMPDHSLSTASKERTHGSEQLHLQGLTVIQNPQTS DSGIYKCTAKNPLGSDYAATYIQVI
2084	A	1	1742	VSAVEFVLHGKDFQVDCKASGSPVP*ISWSLLDGTMI NNAMQADDSGHRTRRYTLFNNGTLYFNKVGVAEEGDY TCYAQNTLGKDEMKVHLTVITAAPRIRQSNKTNKRIK AGDTAVLDCEVTGDPKPKIFWLLPSNDMISFSIDRYT FHANGSLTINKVKLLDSGEYVCVARNPSGDDTKMYKL DVVSKPPLINGLYTNRTVIKATAVRHSHKHFDCAEAG TPSPEVMWIMPDNIFLTAPYYGSRITVHKNGTLEIRN VRLSDSADFICVARNEGGEVSVLVVQLEVLEMLRRPTF RNPFFNEKIVAQLGKSTALNCSVDGNPPPEIIWILPNG TRFSNGPQSYQYLIASNGSFIISKTTREDAGKYRCAA RNKVGYYIEKLVILEIGQKPVILTYAPGTVKGISGESL SLHCVSDGIPKPNIKWTMPSGYVVDPRQINGKYILHD NGTLVIKEATAYDRGNYICKAQNSVGHTLITVPMIV AYPPRITNRPPRSIVTRTGAAFQLHCVALGVPKPEIT

618
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				WEMPDHSLSTASKERTHGSEQLHLQGTTLVIQNPQTS DSGIYKCTAKNPLGSDYAATYIQVI
2085	A	1	1742	VSAVEFVLHGKDFQVDCASGSPVP*ISWSLLDGTMI NNAMQADDSGHRTRRYTLFNNGTLYFNKVGVAEEGDY TCYAQNTLGKDEMKVHLTVITAAPRIRQSNKTNKRK AGDTAVLDCEVTGDPKPKIFWLLPSNDMISFSIDRYT FHANGSLTINKVKLLDSGEYVCVARNPSGDDTKMYKL DVVSKPPLINGLYTNRTVIKATAVRHSHKHFDCAEG TPSPEVMWIMPDNIFLTAPYYGSRTVHKNGTLEIRN VRLSDSADFICVARNEGGEVLLVVQLEVLEMLRRPTF RNPFEKIQAQLGKSTALNCSVDGNPPPEIIWILPNG TRFSNGPQSYQYLIASNGSFIISKTTREDAGKYRCAA RNKVGYIEKLVILEIGQKPVILTYAPGTVKGISGESL SLHCVSDGIPKPNIKWTMPSGYVVDROPQINGKYILHD NGTLVIKEATAYDRGNYICKAQNSVGHTLITVPVMIV AYPPRIITNRPPRSIVTRTGAAFQLHCVALGVKPEIT WEMPDHSLSTASKERTHGSEQLHLQGTTLVIQNPQTS DSGIYKCTAKNPLGSDYAATYIQVI
2086	A	180	275	MEEPQSDPSVEPPLSQETFSDLWKLLSENVL
2087	A	47	1147	MASMAAVLTWALALLSAFSATQARKGFWDYFSQTS KGRVEQIHQQKMAREPATLKDSLEQDLNMMNKFLEKL RPLSGSEAPRLPQDPVGMRRQLQEELEEVEKARLQPYM AEAHELVGWNLEGLRQQLKPYTMDLMEQVALRVQELQ EQLRVVGEDTKAQLLGGVDEAWALLQGLQSRVVHHTG RFKELFHPYAESLVSGIGRHHVQELHRSVAPHAPASPA RLSRCVQVLSRKLTALKAKALHARIQQNLDQLREELSR AFAGTGTEEGAGPDPQMLSEEVQRQLQAFRQDTYLQI AAFTRAIDQETEEVQQQLAPPPPGHSAFAPEFQQTDS GKVL SKLQARLDDLWEDITHSLHDQGHSHLGD*
2088	A	47	1147	MASMAAVLTWALALLSAFSATQARKGFWDYFSQTS KGRVEQIHQQKMAREPATLKDSLEQDLNMMNKFLEKL RPLSGSEAPRLPQDPVGMRRQLQEELEEVEKARLQPYM AEAHELVGWNLEGLRQQLKPYTMDLMEQVALRVQELQ EQLRVVGEDTKAQLLGGVDEAWALLQGLQSRVVHHTG RFKELFHPYAESLVSGIGRHHVQELHRSVAPHAPASPA RLSRCVQVLSRKLTALKAKALHARIQQNLDQLREELSR AFAGTGTEEGAGPDPQMLSEEVQRQLQAFRQDTYLQI AAFTRAIDQETEEVQQQLAPPPPGHSAFAPEFQQTDS GKVL SKLQARLDDLWEDITHSLHDQGHSHLGD*
2089	A	1199	329	DFGEFMRENRLTPFLDPYKIDGSLEVPLERAKDQLE KHTRYWPMIISQTTIFNMQAVVPLASVIVKESLTEED VLNCQKTIYNLVDMERKNDPLPISTVGTRGKGPKRDE QYRIMWNELETLVRAHINNSEKHQRVLECLMACRSKP PEEEERKKRGRKREDKEDKSEKAVKDYEQEKSWQDSE RLKGILERGKEELAEAEI IKDSPDSPEPPNKKPLVEM DETPQVEKSKGPVSLLSLWSNRINTANSRKHQEFAGR LNSVNNRAELYQHLKEENGMEETTENGKASRQ
2090	A	3	456	RWNSIMELALLCGLVVMAGVPIQGGILNLNKMVKQV TGKMPILSYWPYGCHCGLGGRGQPKDATDWCCQTHDC CYDHLKTQGCGLYKDYRYNFSQGNIHCSDKGSWCEQ

619
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QLCACDKEVAFCLKRNLDITYQKRLRFYWRPHCRGQTPGC
2091	A	27	489	EGEPRD*PKMPLTPEPP/VWARGGAPRMGSSPMALT ALWALHPHHAGPGHPCALHPHRCVG*TPVPPCSP RPQPPCTHPGVAPRRRAVD*AHGHRPRAL*GLVWLCG PPADRSGP*ASHPATWAPRPYWRSQPGAPSGGPPSPGR GGPPPQA
2092	A	2022	617	VIPPVLTARGPRPRGAGAMVRGRISRLSVRDVRFPTS LGHHGADAMHTDPDYSAAYVVIETDAEDGIKGCGITF TLGKGTEVVVCAVNALAHHLNLDLKDIDVDFRGFYR QLTSDGQLRWIGPEKGVVHLATAAVLNAVWDLWAKQE GKPVWKLVDMDPRMLVSCIDFRYITDVLTEEDALEI LQKGQIGKKEREKQMLAQGYPAYTTSCAWLGYSDDTL KQLCAQALKDGTWTRFKVKVGADLQDDMRRCQIIRDMI GPEKTLMDANQRWDVPEAVEWMSKLAKFKPLWIEEP TSP*LTFLGHATI\SKALVPFRELGICTRENSCHNRV IFKQLLQAKALQFLQIDSCRLGSVNENLSVLLMAKKF EIPVCPHAGGVGLCELVQHLLIFDYISVSASLENRVC EYVDHLHEHFKYPVMIQRASYMPPKDPGYSTE\LKEE SCKRNTQYPQMGEVWEETPFPAQEN
2093	A	63	193	SGRLAPHTSRRTSANCSDDAKSSDSCSPSRKT*WGR NTNRIH
2094	A	1404	142	IPGSTISWSPAAARGLSVCRCRLHPASAMDLFGDLP EPERSPRPAAGKEAQKGPLLFDLPPASSTDSGSGGP LLFDLPPASSGDSGLATSISQMVKTEGKGAKRKT EEEKNGSEELVEKKVCKASSVIFGLKGYVAERKGERE EMQDAHVILNDITEECRPPSSLI TRVSYFAVFDGHGG IRASKFAAQNLHQNLIKFPKGDVISVEKTVKRCLLD TFKHTDEEFLKQASSQKPAWKDGSTATCVLAVDNILY IANLGDSRAILCRYNEESQKHAALSLSKEHNPTQYEE RMRIQKAGGNVRDGRVLGVLEVSRSIGDGQYKRCGVT SVPDIRRCQLTPNDRFILLACDGLFKVFTPEEAVNFI LSCLEDEKIQTREGKSAADARYEAACNRLANKAVQRG SADNVTVMVVRIGH
2095	A	2	541	FVGHCVNTEGGFVCERGPGMRVSADRHSCQDTDECLG TPCQQRCCKNSIGSYKCSCRTGFHLHGNNHSCV/DYTP RIPLCSPIFLAAFAPLDVNECRRPLERRVCHHSCHNT GGSFLCTCRPGFRLRADRVSC/DFPESRAGPICHPA TPVTPVQE/CYCCLLRPHGLPCAQDIDLLGLQGHQ
2096	A	1206	2266	RHLLTIFHKLKIYKTINKIDFKKKRVTQLLVFCLFLC LFFSSEMVKNTMTVTEFLLLGFLLPRIQMLLFGFLS LFYVFTLLGNGTILGLISLDSRLHTPMYFFLSHLAVV NIAYACNTVPQMLVNLLHPAKPISFAGCMT*TFLFLS FAHTECLLLVLMYSYDRYVAICHPLRYFIIMTWKVCIT LAITSWTCGSLAMVHVSLLRLPFCGPREINHFFCE ILSVLRLACADTWLNQVVI FAACMFILVGPLCLVLVS YSHILAAILRIQS GEGRRKAFSTCSSHLCVVGFLFFGS AIVMYMAPKSRHPEEQQKVLFLFYSSFNPMNLNPLIYN LRNVEVKGALRRALCKESH
2097	A	1206	2266	RHLLTIFHKLKIYKTINKIDFKKKRVTQLLVFCLFLC

620
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LFFSSEMVKNQTMVTEFLLLGFLGPRIQMLLFGFLS LFYVFTLLGNGTILGLISLDSRLHTPMYFFLSHLAVV NIAYACNTVPQMLVNLLHPAKPISFAGCMT*TFLFLS FAHTECLLLVLMYDRYVAICHPLRYFIIMTWKVCIT LAITSWTCGSLAMVHVSLLRLPFCGPREINHFFCE ILSVLRLACADTWLNQVVI FAACMFILVGPLCLVLVS YSHILAAILRIQSGEGRRKAFSTCSSHL CVVGLFFGS AIVMYMAPKSRHP EEQKVLFLFYSSFNPM LNPLIYN LRNVEVKGALRRALCKESH S
2098	A	276	243	EKWPD*SRAACPVL CRGNGQY SKGRCLCFSGWKGT EC DVPTTQCIDPQCGGRGICIMGSCACNSGYKGES CEEAPRYIPEKE
2099	A	4	770	RETGSVSLSPSGLEGAESYAVSPILYSSPDVKELWLE TLQGQRHSHTGVKSTPGQSAAILMKLRSSHNASKTLN ANNMETLIECQSEGD IKEHPLLASCESEDSICQLIEV KKRKKVLSWPFLMRRLSPASDFSGALETDLKASLFDQ PLSIICGDSDTLPRPIQDILTILCLKGPSTEGIFRRA ANEKARKELKEELNSGDAVDLERLPVHLLAVVFKDFL RSIPRKLSSDLFEEWMGALEMQDEEDRIEALK
2100	A	901	521	FFFGNGVSPCRQAGV*WHDLDLQNLPPGFKRFSYLS LPSSW\DYRHVLPRQANFCIF/M*RRGFTMLARMVSI S*PRDLPALASQSAGITGVSHHAPPQMDFTFALLCFA LKGCLPRQKEGGTLNLI
2101	A	901	521	FFFGNGVSPCRQAGV*WHDLDLQNLPPGFKRFSYLS LPSSW\DYRHVLPRQANFCIF/M*RRGFTMLARMVSI S*PRDLPALASQSAGITGVSHHAPPQMDFTFALLCFA LKGCLPRQKEGGTLNLI
2102	A	3	600	PRCRNSARVADTFYTNAGCTLVALNPFPKVPQLYSPE LMREYHAAPQPQKLKPHVFTVGEQTYRNVKSLIEPVN QSIVVSGESGAGKTWTSRCLMKFYAVVATSPASWESH KIAERIEQRI LNSNPVMEAFGNACTLRNNNSSRFGKF IQLQLNRAQQMTGA AVQTYLLEKTRVACQASSERNKD PIPPELTRLLQQSQ
2103	A	3	600	PRCRNSARVADTFYTNAGCTLVALNPFPKVPQLYSPE LMREYHAAPQPQKLKPHVFTVGEQTYRNVKSLIEPVN QSIVVSGESGAGKTWTSRCLMKFYAVVATSPASWESH KIAERIEQRI LNSNPVMEAFGNACTLRNNNSSRFGKF IQLQLNRAQQMTGA AVQTYLLEKTRVACQASSERNKD PIPPELTRLLQQSQ
2104	A	10	435	FKWLLKSHAICFWTRS*SYCDNVCVPSLWAHHLGIRT EIP EFFLSKFLCTSIIPHFTYRRQLRLIQGST*EA* EDKLEQK*ALGAAQFTLPGMDVFVCFVFCF/CLFEME SHSVT*ARVQWCDLGS LQPLPLGFKQFSC LGL
2105	A	79	1222	CQRREDAEFWLCFALDPSKDPCLKVKCSPHKVCVTQ DYQTALCVSRKHL LPRQKKNVAQKHWVGPSNLVKCK PCPVAQSAMVCGSDGHSYTSKCKLEFHACSTGKSLAT LCDG\PCPCLPEP\EPPKHKGKRGVPCTDKELRNLAS RLKDWFGALHEDANRVIKPTSSNTAQGRFDTSILPIC KDSL GWMLNKLD MN YD LLLDPSEINAIYLDKYEP CIK PLFNSCDSFKDGKPF LNNEWCLLP SQNPGLP/CAQN

621
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EMNRIQ\KLSKGKSLLGAFIPRCNEEGYYKATQCHGS TGQCWCVDKYGNELAGSRKQGAVSCEEEQETSGDFGS GGSVLLDDLEYERELGPKDKEGKLRVHTRAVTEDDE DEDDDKEDDEVGYIW
2106	A	174	857	MLNLAFTVGSFLLSAITLPLGIVMDKYGPRKLRLLS ACFAVSCLLIAYGASKPNALSVLIFIALALNGFGGMC MTFTSLTLPNMFGLRFTFIALMIGSYASSAVTFPGI KLIYDAGVSFIVVLVWAGCSGLVFLNCFNWPPEPF PGPEDMDYSVKIKFSWLGFHDHKITGKQFYKQVTTVGR RLSVGSSMRSKEQVALQEGHKLCLSTVDRNSXRSXA LVSGYP
2107	A	174	857	MLNLAFTVGSFLLSAITLPLGIVMDKYGPRKLRLLS ACFAVSCLLIAYGASKPNALSVLIFIALALNGFGGMC MTFTSLTLPNMFGLRFTFIALMIGSYASSAVTFPGI KLIYDAGVSFIVVLVWAGCSGLVFLNCFNWPPEPF PGPEDMDYSVKIKFSWLGFHDHKITGKQFYKQVTTVGR RLSVGSSMRSKEQVALQEGHKLCLSTVDRNSXRSXA LVSGYP
2108	A	1	570	YAAFGAVVTRVSLPAPRCPALGGLASGPGESGPALLQ VCGAKCPGGAPRGENREKEETTRIGPGVMESKEKRAV NSLSMENANQENEEKEQVANKGEPLALPLDAGEYCVF RGNRRRFRVRQPILQYRWDMHRLGEPQARMREENME RIGEEVRQLMEKLREKQLSHSLRAVSTDPHHDHDE FCLMP
2109	A	70	993	SEQKIQEQGYVWITVFSALPTTVSALHPRVLKPLSSL IHLQANSNPWECNCKLLGLRDWLASSAITLNIYWQNP PSMRGRALRYINITNCVTSSINVSRAWAVVKSPHIIH KTTALMMAWHKVTTNGSPLENTETENITFWERIPTSP AGRFFQENAFGNPLETTAVLPVQIQLTTSVTNLNLEKN SALPNDAAASMSGKTSLICTEVEKLEAFDILLAFFI LACVLIIFLIYKVQFKQKLKASENSRENREYYSFY QSARYNVTASICNTSPNSLES PGLEQIRLHKQIVPEN EAQVILFEHSAL
2110	C	160	297	MILCHLMQAPYHLKVSWEPTDPPTLWKCWTNVSTNPP LSALRGHR
2111	A	2	951	PRVRPRVRPRVRSSRPRSRDPSRRARLRWQLRWKPR WCPRPPKTPGVWKRPRTRPRSSAGGSTGFPSSPILRR SPSTRRRSSRKASPTATRATGTPPRQAQRKTARAAGR RRASPGIATAGTRSMISM\RPGRKPSNPSWEGRTNEE TSSLSRLKPVSPGTITCPLRTPGSLKDSKIPISIKH LTNLPSSHPVVHQQPSRSEMPRTKIPVSKVLVRRVSN RGLAGTTIRATACHDSAQKVVRSSRPRWMGMPRNTT FPWETTKVSFAFPKESLL/WTPPVPRPAPERGPRRSL CPE*GPDNTRKRDATRGFLLSR
2112	A	82	435	MLVLLPRSKAMPLLSVNVTLAFFPRNKEIVKYLLNQG ADVTLRAKNGYTAFDLVMLLNDPDI FG GELIGFLSVV TELVRLLASVFMQVNKDIGRRSHQLPLPHSKVPTALE HPSAAR*
2113	A	83	1138	PRRMGSWVQLITSVG VQQNH PGWTVAGQFQEKKRFE EVIEYFQKKVSPVHLKILLTSDEAWKRFVRVAELPRE

622
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				EADALYEALKNLTPTYVAIEDKDMQQKEQQFREWFLKE FPQIRWKIQESI ERLRVIANEIEKVHRGCVIANVVSG STGILSVIGVMLAPFTAGLSLSITAAGVGLGIASATA GIASSIVENTYTRSAELTASRLTATSTDQLEALRDIL HDITPNVLSFALDFDEATKMIANDVHTLRRSKATVGR PLIAWRYVPINVVETLRTGAPTRIVRKVARNLGKAT SGVLVLDVNVNLVQDSL DLHKGEKSESAELLRQWAQE LEENLNELTHIHQSLKAG
2114	A	83	1138	PRRMGSWVQLITSVGVQONHPGWTVAGQFQEKKRFT EVIEYFQKKVSPVHLKILLTSDEAWKRFVRVAELPRE EADALYEALKNLTPTYVAIEDKDMQQKEQQFREWFLKE FPQIRWKIQESI ERLRVIANEIEKVHRGCVIANVVSG STGILSVIGVMLAPFTAGLSLSITAAGVGLGIASATA GIASSIVENTYTRSAELTASRLTATSTDQLEALRDIL HDITPNVLSFALDFDEATKMIANDVHTLRRSKATVGR PLIAWRYVPINVVETLRTGAPTRIVRKVARNLGKAT SGVLVLDVNVNLVQDSL DLHKGEKSESAELLRQWAQE LEENLNELTHIHQSLKAG
2115	A	700	283	VPRLVSPLSNPAPKFYCVSFFYHMYGKHIGSLNLLVR SRNKGALDTHAWSLSGNKGNVWQQAHVPI SPSGPFQI IFEGVRGPGYLGDI AIDDVTLKKGECPRKQTDPNKVV VMPGSGAPCQSSPQLWGPM AIFLLALQR
2116	A	700	283	VPRLVSPLSNPAPKFYCVSFFYHMYGKHIGSLNLLVR SRNKGALDTHAWSLSGNKGNVWQQAHVPI SPSGPFQI IFEGVRGPGYLGDI AIDDVTLKKGECPRKQTDPNKVV VMPGSGAPCQSSPQLWGPM AIFLLALQR
2117	A	554	970	MVLPPFICNLLRRHPACRVLVHRPHGPELDADPYDPGE EDPAQSRALESSLWELQALQRHYHPEVSKAASVINQA LSMPEVSIAPLLELTAYEIFERDLKKKGPEPVPTGVL SQPRACWDGRVKLCAQHFAQLTLAHL*
2118	A	1	541	VHVCSSKMGALSTERLQYYTQELGVRERSGHSVSLID LWGLLVEYLLYQEENPAKLS DQQA EAVRQGQNPYPIYT SVNVRTNLSGEDFAEWCEFTPYEVGF PKYGAYVPTTEL FGSELFMGRLLQLQPEPRICYLQGMWGS AFATSLDEI FLKTAGSGLSFLEWYRGSVNITDDCQKPQLHN
2119	A	1	541	VHVCSSKMGALSTERLQYYTQELGVRERSGHSVSLID LWGLLVEYLLYQEENPAKLS DQQA EAVRQGQNPYPIYT SVNVRTNLSGEDFAEWCEFTPYEVGF PKYGAYVPTTEL FGSELFMGRLLQLQPEPRICYLQGMWGS AFATSLDEI FLKTAGSGLSFLEWYRGSVNITDDCQKPQLHN
2120	A	1	1524	PHPSGPRITHSHARETACQP/GSEQHPGPHGGQLPRG GRQGPELP SHVCRAQA\GRTGQEPSSERPHAGQGAGL WSGSPWGRGRTQPTHAPTEGATPRCPLRPSPRGSGRA GPTLIRAGLSGGRGGRSLCPCGFPRAGAVPARSSHNQ TSPVHEKSRH/GPTASGPGCW L GDPQGRVPGLAVP *APAACTPMDKLPGLHLPEQRLPSIGGPF SAGLSPSG QSREWQGSQGSRSRQFSKKAPGPPPS\TGGGCLGCG GRGT\RGSAHAG\PWGSPHQQS*GAPGSQAKGGTP* RKPA PANGSSEEQEARGPQGLEVSSSQT SASHAGLG LQGNSTRGVGPGPRPPAEPTTGRSWARSRVNPD*EQA

623
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				SGA*VRSGSRSPGDALESSCNAPAWLQLCSAPCALGS REPGQGLAVTQTLGPGQSLGHPRESHKTRPRYEAATS SACLGALALTGTFSVEETEMFMTRQRPTGRDLQRGTRP QGWQGPVPGTSHYGRARPALGEASDKQEANGA
2121	A	233	692	DNHPSFPRLPSSRPGTKEVLKEIHIISDTTADVIFYPI YRMSEMIFFRIKMPWLWLDLWYLMFKEGWHEHKKSLKI LHTFTNSVIAERANEMNANEDCRGDGRGSAPSKNKRR AFLDLLLSVTDDEGNRLSHEDIREEVDTFMFEVLYIV RFRYH
2122	A	2	1115	PRVRSSGGQEDPASQQWARPRFTQPSKMRRRVIARPV GSSVRLKCVASGHPRPDI TWMKDDQALTRPEAAEPRK KKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDV IQRTSRKPVLTGTHPVNTTVDFGGTTSFQCKVRSQDV PVIQWLKRVEYGAEGRHNSIDVGGQKFVVLPTGQDVW SRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFR SAFLTVPDPKPPGPPVASSSSATSLPWPVVIGIPAG AVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPPGTA RDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLL GPGPVAGPKLYPKLYTGHSTPHTYTHPPSPCQLNSSH S
2123	A	2	1115	PRVRSSGGQEDPASQQWARPRFTQPSKMRRRVIARPV GSSVRLKCVASGHPRPDI TWMKDDQALTRPEAAEPRK KKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDV IQRTSRKPVLTGTHPVNTTVDFGGTTSFQCKVRSQDV PVIQWLKRVEYGAEGRHNSIDVGGQKFVVLPTGQDVW SRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFR SAFLTVPDPKPPGPPVASSSSATSLPWPVVIGIPAG AVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPPGTA RDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLL GPGPVAGPKLYPKLYTGHSTPHTYTHPPSPCQLNSSH S
2124	A	2	1115	PRVRSSGGQEDPASQQWARPRFTQPSKMRRRVIARPV GSSVRLKCVASGHPRPDI TWMKDDQALTRPEAAEPRK KKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDV IQRTSRKPVLTGTHPVNTTVDFGGTTSFQCKVRSQDV PVIQWLKRVEYGAEGRHNSIDVGGQKFVVLPTGQDVW SRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFR SAFLTVPDPKPPGPPVASSSSATSLPWPVVIGIPAG AVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPPGTA RDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLL GPGPVAGPKLYPKLYTGHSTPHTYTHPPSPCQLNSSH S
2125	A	3	644	PNWKRNP SLF*KVF PFMKKVV/QRGSLLPPKSLDYDR FSRN/DTPLGRVSIPLNKVDLTQMOTFWKDLKPCSDG SGSRGELL LSLCYNPSANSIIVNIIKARNLKAM\DIG GTSDP\YVKVWL\MYK\DKRV\EKKKTVT\MKRNLP \IFNESFAFDI PTEKLRETTIIITVMDKDKLSRNDVI GKIYLSWKSGPGEVKHWKDMIARPRQPVAQWHQLKA
2126	A	193	883	IMPCAQRSWLANLSVVAQLLNFGALCYGRQPQPGPVR FPDRRQEHFIKGLPEYHVVG PVRVDASGHFSLSYGLHY

624
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PITSSRRKRDLGSEDWVYYRIFHEEKDLFFNLTVNQ GFLSNSYIMEKRYGNLSHVKMMASAPLCHLSGTVLQ QGTRVGTAAALSACHGLTGFFQLPHGDFEIEPVKKHPL VEGGYHPHIVYRRQKVPETKEPTCGLKGI VTHMSSWV EESVLFFW
2127	A	87	477	IKSKLNQQVEVQSEWRLTEAKGPTMGKESGWDSGRA AVAAVGGVVAVGTVLVALSAMGFTSVGIAASSIAAK MMSTAAIANGGGVAAGSLVAILQSVGAAGLSVTISKVI GGFAGTALGAWLGSPSS
2128	A	1993	1379	SLHLSERADWQYSQRAG/DAVEVFFSRTARDNRLGCM FVRCAPSSRYTLLFSHGNAVDLGQMCSFYIIGLSRIN CNIFSYDYSYGVSSGKPKSEKNLYADIDAAWQALRTR YGVSPENIILYGQSIGTVPTVDLASRYECAAVILHSP LMSGRLRVAFPDTRKTYCFDAFPSIDKISKVTSPVLVI HGTEDEVIDFSHGLAMYERCPRAVEPLWVEGAGHNDI ELYAQYLERLQKFISHEL PNS *RQSK
2129	A	1993	1379	SLHLSERADWQYSQRAG/DAVEVFFSRTARDNRLGCM FVRCAPSSRYTLLFSHGNAVDLGQMCSFYIIGLSRIN CNIFSYDYSYGVSSGKPKSEKNLYADIDAAWQALRTR YGVSPENIILYGQSIGTVPTVDLASRYECAAVILHSP LMSGRLRVAFPDTRKTYCFDAFPSIDKISKVTSPVLVI HGTEDEVIDFSHGLAMYERCPRAVEPLWVEGAGHNDI ELYAQYLERLQKFISHEL PNS *RQSK
2130	A	3	383	PPGPKGDQGDGEGKEGRPGIPGLPGLRGLPGERGTPGL PGPKGNDGKLGATGPMGMRGFKGDRGPKGEKGEKGR AGDASGVEAPMMIRLVNNGSGPHEGRVEVYHRRWGTV CDDGWDKKDGDVVC RM
2131	A	3	383	PPGPKGDQGDGEGKEGRPGIPGLPGLRGLPGERGTPGL PGPKGNDGKLGATGPMGMRGFKGDRGPKGEKGEKGR AGDASGVEAPMMIRLVNNGSGPHEGRVEVYHRRWGTV CDDGWDKKDGDVVC RM
2132	A	1	2789	GIRTSSPKTEGKHEETVNKESDMKVPTVSLKVSESVI DVKTTMESISNTSTQSLTAETKDIALEPKEQKHEDRQ SNTPSPPVSTFSSGTSTTSDIEVL DHESVISESSASS RQETTDSSKSSLHLMQTSFQLLSASACPEYNRLDDFQK LTESCCSSDAFERIDSFSVQSLDSRSVSEINSDDELS GKGYALVPIIVNSSTPKSKTVESAEGKSEEVNETLVI PTEEAEMEESGRSATPVNCEQPDILVSSTPINEGQTV LDKVAEQCEPAESQPEALSEKEDVCKTVEFLNEKLEK REAQLLSLSKEKALLEEAFDNLKDEMFRVKEESSIS SLKDEFTQRIAEAEKKVQLACKERDAAKKEIKNIKEE LATRLNSSETADLLKEKDEQIRGLMEEGEKLSKQQLH NSNIKKLRAKDKENENMVAKLNKKVKELEELQHLK QVLDGKEEVEKQHRENIKKLNSMVERQE KDLGR LQVD MDELEEKNRSIQAALDSAYKELTDLHKANA AKDSEAQ EAALSREMKAKEELSAALEKAQEEARQQQETLAIQVG DLRLALQRTEQAARKEDYLRHEIGELQQR LQEAENR NQELSQSVSSTTRPLL RQIENLQATLG SQTSSWEKLE KNLSDRLGESQTLLAAVERERAATEELLANKIQMSS MESQNSLLRQENS RFQAQLESEKNRLCKLE DENNRYQ

625
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VELENLKDEYVRTLEETRKEKTLLNSQLEMERMKVEQ ERKKAI FTQETIKEKERKPFVSSTPTMSRSSSISGV DMAGLQTSFLSQDESHDSFGMPIS/AKWKQSL*CC KDGSRIKH\ IENLQSQLKLREGEITHLQLEIGNLEKT RSIMAEELVKLTNQNDELEEKVKEIPKLRTQLRDLDO RYNTILQMYGEKAEAEELRLDLEDVKNMYKTQIDEL LRQSLS
2133	A	1	2234	MAASSIRDERTRTYYLPVVRAPYTCNFRPSSAAVGR GGWGRAQKWNNSGKCRFWEVSESLTLEDVAVEFTWEE WQLLGPAQKDLYRDVMLENYSNLVSVGYQASKPDALF KLEQGEPWTVENEIHSQICPGMYALYRKKHNGYRVKY DSEFQASMVGVSWNISPIDEGLLYIYKRHKEFTTEV DKGCETNIOMKDDKI KKVDNHLQMHSSQKQRCLKRVEQ CHKHNAFGNIIHQKSDFLRQNHDTFDLHGKILKSN LSLVNQNKRYEIKNSVG VNGDGKSFLHAKHEQFHNM NFPEGNSVNTNSQFIKHQRTQNIDKPHVCTECGKAF LKKSRLIYHQRVHTGEKPHGCSICGKAFSRKSGLTEH QRNHTGEKPYECTECDKA FRWKSQ LNAHQKIHTGEKS YICSDCGKGFIIKKSRLINHQRVHTGEKPHGCSLCGKA FSKRSRLTEHQRTHTGEKPYECTECDKA FRWKSQ LNA HQAHTGEKSYICRDCGKGF IQGNLIVHQRIHTGEK PYICNECGKGF IQGNLLIHRRTHTGEKPYVCNECGK GFSQKTCLISHQRFHTGKTPFVCTECGKSCSHKSLI NHQRIHTGEKPYTCSDCGKA FRDKSCLNRHRRHTHTGE RPYGCSDCGKA FSHLSCLVYHKGMLHAREKCVG/CSQ IGKSLLRES*LITYT*SHTG*RLC*HGDSADAFCGSS DLIN*QCVPSREQSSHCEPACCQKFSLSR**NCHGIK NHYECR
2134	A	3	713	RLAFPCGRPDYWALARRTIGTGLERKALGLPGSSERP TSVSSYQGTIRIRCSNPGGKMRPLTEEETRVMFEEKIAK YIGENLQLLVDRPDGTYCFRLHNDRVYYVSEKIMKLA ANISGDKLVSLGTCFGKFTKTHKFR LHVTALDYLAPY AKYKVIKPGAEQSFLYGNHVLKSGLG RITENTSQYQ GVVVYSMADIPLGFGVAAKSTQDCRKVDPMIIVFHQ ADIGEYVRHEETLT
2135	A	1	350	EGGTGVRSLSFYQHIITVGTGHGSLLFYDIRAQKFLE ERASSLD SMPGPAGRKLKLACGRGWL NQDDVWVNYF GGMGEFPNALYTHCYNWPEMKLFVAGGPLPSGLHGNY AGLWS
2136	B	238	1323	XESVEIVSEVRVEVGELNIIKDWGRESVEKGGAVISM EAERVKGQAMIATGGVITGLAALKRQDSARSQQHVNL SPSPATQEKKPIRRRPRADV VVRGKIRLYSPSGFFL ILGVLSIIIGIAMAVLGYWPQKEHFIDAETTLSTNET QVIRNEGGVVVRFFEQLHSDKMKMLGPFTMGIGIFI FICANAILHENRDKETKIIHMRDIYSTVIDIHTLRIK EQRQMNGMYTGLMGETEVKQNGSSCASRLAANTIASF SGFRSSFRMDSSVEEDELMLNESKSSGHLMPPLSDS SVSVFGLYPPPSKTTDDKTSGSKK CETKSIVSSSISA FTLPVIKLNNCVIDEPSIDNITEDADNLKX
2137	A	41	1285	VGEMTLIWRHLLRPLCLVTSAPRILEMHPFLSLGTSR

626
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TSVTKLSLHTKPRMPPCDFMPERYQVIFLVNSGSEAN ELAML MARAHSN NIDII SFRGAYHGCSPYTLGLTNVG IYKMELPGGTGCQPTMCPDVFRGPWGGSHCRDSPVQT IRKCSCAPDCCQAKDQYIEQFKDTLSTSVAKSIAGFF AEPIQGVNGVVQYPKGFLKEAFELVRARGGVCIANEV QTGFGR LGSHFWGFQTHDVL PDI VTMAGGIGNGF PMA AVITTP EIAKSLAKCLQHFN TFGGNPMACAIGSAVLE VIKEENLQENSQEVGT YM L LKFAKL RDEF EIVGDVRG KGLMIGIEMVQDKI SCRPLPREEVNQIHEDCKHMGLL VGRGSIFSQT FRIAPSMCITKPEVDFAVEVFRSALTQ HMERRAK
2138	A	41	1285	VGEMTLIWRHLLRPLCLVTSAPRILEMHPFLSLGTSR TSVTKLSLHTKPRMPPCDFMPERYQVIFLVNSGSEAN ELAML MARAHSN NIDII SFRGAYHGCSPYTLGLTNVG IYKMELPGGTGCQPTMCPDVFRGPWGGSHCRDSPVQT IRKCSCAPDCCQAKDQYIEQFKDTLSTSVAKSIAGFF AEPIQGVNGVVQYPKGFLKEAFELVRARGGVCIANEV QTGFGR LGSHFWGFQTHDVL PDI VTMAGGIGNGF PMA AVITTP EIAKSLAKCLQHFN TFGGNPMACAIGSAVLE VIKEENLQENSQEVGT YM L LKFAKL RDEF EIVGDVRG KGLMIGIEMVQDKI SCRPLPREEVNQIHEDCKHMGLL VGRGSIFSQT FRIAPSMCITKPEVDFAVEVFRSALTQ HMERRAK
2139	A	3	362	EGKPASAI VGGK PANILEFPWHV GIMNHGSHLCGGSI LNEWVLSASHCFDQLNNSKLEI IHGTEDLSTKGIKY QKVDKLF LHPKFDDWLLDNDIALLLLKSPLNLSVNRI PICTSEISD
2140	A	1	663	EIANLILAENCEAALALHLYRGGRLLOGHRIPFGVIF GGTDVNEDANQAEKNTVMGRVLEEAFVAFTESMKE MAQAQWVDPVFTREV KAKVKRAAGVRLIGEMPQEDLH AVVKNCFAVVNSSVSEGMSAAILEAMDLEVPVLARNI PGNAAVVKHEVTGLLFSNPQEFVHLAKRLVSDPALEK EIVVNGREYVRMYHSWQVERDTYQQLIRKLEGSTED
2141	A	8	1516	MSLVLLSLAALCRSAVPREPTVQCGSETGPSPEWMLQ HDLIPGDLRLRVEPVTT SVATGDYSILMNVS WVLRA DASIRLLKATKICVTGKS NFQSYSCVRCNYTEAFQTQ TRPSGGKWTFSYIGFPVELNTVYFIGAHNIPNANMNE DGPSMSVNFTSPGCLDHIMKYKKKCVKAGSLWDPNIT ACKKNEETVEVNFTTTPLGNRYMALIQHSTIIGFSQV FEPHQKKQTRASVVI PVTGDSEGATVQLTPYFPTCGS DCIRHKGT VVLC PQTGVPFPLDNNKSKPGGWLPLLLL SLLVATWVLVAGIYLMWRHERIKKTSFSTTTLLPPIK VLVVYPSEICFHHTICYFTEFLQNHCRSEVILEKWQK KKIAEMGPVQWLATQKKAADKVVFLLSNDVNSVCDGT CGKSEGPSSENSQDLFPLAFNLFCSDLRSQIHLHKYV VYFREIDTKDDYNALSVCPKYHLMKDATAFCAELLH VKQQVSAGKRSQACHDGCCSL*
2142	A	1	622	PDPCLNNGGSCVDLVGN YTC LCAEPFKGLRCETGDHPV PDACLSAPCHNGGTCVDADQGYVCEYPEGFMGLDCRE RVPDDCECRNGGRCLGANTTLCQCPLGFFGLLCEFEI

627
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TAMPCNMNTQCPDGGGYCMEHGGSYLCQCPLGFFGLLC EFEITAMPCNMNTQCPDGGGYCMEHGGSYLCVCHTDHN ASHSLPSPCSDSPSSCQAWRNE
2143	A	3	87	PAMNVNDSVTKQKFDNLYCCRESILDG
2144	A	406	888	IASQNFDPATVSVATAHKGAEPSRGTAWGPVAKRLQQ ELMTLMMPGDKRISAYPESLIKWTPSMKQLAQCMKI* GISSCWSSSMTTHLYDAPTVKFLTPCYHPNVDTOGNI CLDILKEKWSAPYDIRTILLSIQCLLGQLNIDSPLNT HATKLWENPIALR
2145	A	46	1576	APYLPDPMKHTLALLAPLLGLGLGLALSQLAAGATDC KFLGPAEHLTFTPAARARWLAPRVRAPGLLDSLYGTV RRFLSVVQLNPFPSSELVKALLNELASVKVNEVVRYEA GYVVCVIAAGLYLLLVPVPTAGLCFCCCRCHRRCCGRVK TEHKALACERAALMVFLLLTLLLLIGVVCFAVTNQR THEQMGPSIEAMPETLLSLWGLVSDVPQ/GVGVSIGS AIHTQLRSSV\TPCLAAVGS LGQVLQVSVHHLQTLNA TVVELQAGQQDLEPAIREHRDRLELLQE/SQVPSVD HVLHQLKGVPEANFSSMVQEENSTFNALPALAAMQTS SVVQELKKAVAQQPEGVRTLAEGFPGLEAASRWAQAL QEVVESSRPYLQEVQRYETYRWIVGCVLCSVVLFFVL CNLLGLNLGIWGLSARDDPSHPEAKGEAGARFLMAGV GLSFLFAAPLILLVFATFLVGGNVQTLVCQSWENGEL FEFADTPGNLPPSMNLSQLLGLRKNISIHQAY
2146	A	3	717	DLKDTIGSVTKTPSGLYIIHPEGSSYPFEVMCDMDYR GGGWTVIQKRIDGIIDFQRLWCDYLDGFGDLLGEFWL GLKKIFYIVNQKNTSFMLYVALESEDDTLAYASYDNF WLEDETRFFKMHLGRYSGNAGDAFRGLKKEDNQAMP FSTSDVDNDGCRPACLVNGQSVKSCSHLHNKTGWWFN ECGLANLNGIHHFSGKLLATGIQWGTWTKNNSPVKIK SVSMKIRRMYPYFK
2147	A	3	717	DLKDTIGSVTKTPSGLYIIHPEGSSYPFEVMCDMDYR GGGWTVIQKRIDGIIDFQRLWCDYLDGFGDLLGEFWL GLKKIFYIVNQKNTSFMLYVALESEDDTLAYASYDNF WLEDETRFFKMHLGRYSGNAGDAFRGLKKEDNQAMP FSTSDVDNDGCRPACLVNGQSVKSCSHLHNKTGWWFN ECGLANLNGIHHFSGKLLATGIQWGTWTKNNSPVKIK SVSMKIRRMYPYFK
2148	A	3	717	DLKDTIGSVTKTPSGLYIIHPEGSSYPFEVMCDMDYR GGGWTVIQKRIDGIIDFQRLWCDYLDGFGDLLGEFWL GLKKIFYIVNQKNTSFMLYVALESEDDTLAYASYDNF WLEDETRFFKMHLGRYSGNAGDAFRGLKKEDNQAMP FSTSDVDNDGCRPACLVNGQSVKSCSHLHNKTGWWFN ECGLANLNGIHHFSGKLLATGIQWGTWTKNNSPVKIK SVSMKIRRMYPYFK
2149	A	1397	1565	DRLESLLMHIPGVYPNQWNTNFYLFIFYFEAESHVA QTGLQ*RHLGSLQLPPPQV
2150	A	836	633	MSRNLRTALIFGGFISLIGAAFYPIYFRPLMRLEEYK KEQAINRAGIVQEDVQPPGLKVWSDPFGGRK*
2151	A	294	1568	MSLTIWTVCGVLSLFGALS Y AELGTTIKKSGGHYTYI LEVFGPLPAFVRVWVELLIIRPAATAVISLAFGRYIL

628
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EPFFIQCEIPELAIKLITAVGITVVMVLNSMSVSWSA RIQIFLTFCKLTAILIIIVPGVMQLIKGQTQNFKDAF SGRDSSITRLPLAFYYGMYAYAGWFYLNFTVEEVENP EKTIPLAICISMAIVTIGYVLTNVAYFTTINAEELLL SNAVAVTFSERLLGNFSLAVPIFVALSCFGSMNGGVF AVSRLFYVASREGHLPEILSMIHVRKHTPLPAVIVLH PLTMIMLFSGDLDSSLNFLSFARWLFIGLAVAGLIYL RYKCPDMHRPFKVPLFIPALFSFTCLFMVALSLYSDP FSTGIGFVITLTGVPAYYLFIIWDKKPRWFRIMSEKI TRTLQIILEVVPEEDKL*
2152	A	217	378	KNLFYSLSLICSSYP SILDHIVHIIELIGRIPRRFSL SGKYSQDFFSHRGSIVM
2153	A	2046	4541	MTLALAYLLALPQVLDANRCFEKQSPSALSLOLAAYY YSLQIYARLAPCFRDKCHPLYRADPKELIKMVTTRHVT RHEHEAWPEDLISLTKQLHCYNERLLDFTQAQILQGL RKGVDVQRFRTADDQYKRETILGLAETLEESVYSIAIS LAQRYSVSRWEVFMTHLEFLFTDSGLSTLEIENRAQD LHLFETLKTDPFAFHQHMVKYIYPTIGGFDHERLQYY FTLLENCGCADLGNCAIKPETHIRLLKKFKVVASGLN YKKLTDENMSPLEALEPVLSSQNILSISKLVPKIPEK DGQMLSPSSLYTIWLQKLFWTGDPHLIKQVPGSSPEW LHAYDVCMKYFDRLHPGDLITVVDAVTFSPKAVTKLS VEARKEMTRKAIKTVKHFIEKPRKRNSEDEAQEAKDS KVTYADTLNHLEKSLAHLETLSHSFILSLKNSEQETL QKYSHLYDLRSRSEKEKLHDEAVAI CLDGQPLAMIQQL LEVAVGPLDISPKDIVQSAIMKII SALS GGSADLGGP RDPLKVLEGVVA AVHASVDKGEELVSPEDLLEWLRPF CADDAPVVRPRIHV LQILGQS FHLTEEDSKLLVFFRT EAILKASWPQRQVDIADIENEENRYCLFMELLESSH EAEFQHLVLLLQAWPPMKSEYVITNNPWVRLATVMLT RCTMENKEGLGNEVLKMCRLYNTKQMLPAEGVKELC LLLLNQSLLLPSLKL LLESRDEHLHEMALEQITAVTT VNDSNCDQELLSLLLDAKLLVKCVSTPFYPRIVDHL ASLQQGRWDAEELGRHLREAGHEAEAGSLLLA VRGTH QAFRTFSTALRAAQHWV*
2154	A	2046	4541	MTLALAYLLALPQVLDANRCFEKQSPSALSLOLAAYY YSLQIYARLAPCFRDKCHPLYRADPKELIKMVTTRHVT RHEHEAWPEDLISLTKQLHCYNERLLDFTQAQILQGL RKGVDVQRFRTADDQYKRETILGLAETLEESVYSIAIS LAQRYSVSRWEVFMTHLEFLFTDSGLSTLEIENRAQD LHLFETLKTDPFAFHQHMVKYIYPTIGGFDHERLQYY FTLLENCGCADLGNCAIKPETHIRLLKKFKVVASGLN YKKLTDENMSPLEALEPVLSSQNILSISKLVPKIPEK DGQMLSPSSLYTIWLQKLFWTGDPHLIKQVPGSSPEW LHAYDVCMKYFDRLHPGDLITVVDAVTFSPKAVTKLS VEARKEMTRKAIKTVKHFIEKPRKRNSEDEAQEAKDS KVTYADTLNHLEKSLAHLETLSHSFILSLKNSEQETL QKYSHLYDLRSRSEKEKLHDEAVAI CLDGQPLAMIQQL LEVAVGPLDISPKDIVQSAIMKII SALS GGSADLGGP RDPLKVLEGVVA AVHASVDKGEELVSPEDLLEWLRPF

629
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				CADDAWPVRPRIHVLQILGQSFHLTEEDSKLLVFFRT EAILKASWPQRQVDIADIENEENRYCLFMELLESSH EAEFQHLVLLLQAWPPMKSEYVITNNPWVRLATVMLT RCTMENKEGLGNEVLKMCRLYNTKQMLPAEGVKELC LLLLNQSLLLPSLKLLLESRDEHLHEMALEQITAVTT VNDSNCDQELLSLLLDKALLVKCVSTPFYPRIVDHL ASLQQGRWDAEELGRHLREAGHEAEAGSLLLAVRGTH QAFRTFSTALRAAQHWV*
2155	A	2	362	QELERSMAQRCVCVLALVAMLLLVFPTVSRSMGPRSG EHQRASRIPSQFSKEERVAMKEALKVFPTVVSTSFIO HEVVEEYSHLFTIQGSDPSLQPYLLMAHFDVVPAPEE GWEVPPFSG
2156	A	940	2040	MALRFLGFLLAGVDLGVYLMRLELCDPTQRLRVALA GELVGVGGHFLFLGLALVSKDWRFLQRMITAPCILFL FYGWPGLFLESARWLIVKRQIEEAQSVLRILAERNRP HGQMLGEEAQEALQDLENTCPLPATSSFSFASLLNYR NIWKNLLILGFTNFIAHAIRHCYQPVGGGGSPSDFYL CSLLASGTAALACVFLGVTVDVFRGRRGILLLSMTLTG IASLVLLGLWDYLNAAITTFSVLGLFSSQAAAILST LLAAEVIPTTVRGRGLLIMALGALGGLSGPAQRLHM GHGAFLQHVLAACALLCILSIMLLPETKRKLLPEVL RDGELCRRPSLLRQPPPTRCDHVPLLATPNPAL*
2157	A	317	3	MYALLGVFCLAILVFLINCATFALKYRHKQVPLEGQA SMTHSHDWVWLGNEAELLESMDAPPQDEHTTTIDR GPGACEESNHLNNGGSHKHVQSQIHRADS
2158	A	3	1048	LLRARSPOGSEAGVGGAYMLSKGWWKEGRHGGHRRP RGWGAAGRRQSVPGGPAAP/PCTLYSVGADGRGQGHQ SRGCRPPGPPSASSAPCLAWGAAGRARRREG/RSGRCR TEFSPGCTRR*ALT\CGAGPCRR*SR*RGTRRCLRPW ASPGTGAACGRCCCPP*PHLFWLPPSLRLPAEMLLA GSRPTPACRSSPGGSVHTTTGSPASRRGSRGRSRP SPRPRPSVLSCHGVSL*TGRGRRRGCPRARGRA/GV APPSCRKSAR\CGGRPALRRAGPPSCALGPGAPPPI WAPETAEPAPAVPCPERPGCPAPAAAPRPLSPDPAQL PALARLRPSPGFGERAHAQPA
2159	A	190	2392	VPGECDGITSMASGPGTRLRLNLPVMGDGLETSQM STTQAQAQPPANAASTNPPPPETSNNPKPKRQTNQL QYLLRVVLKTLWKHQFAWPFQQPVDVAVKLNLPDYKI IKTPMDMGTIKKRLNNYYWNAQECIQDENTMFTNCY IYNKPGDDIVLMAEAELEKFLQKINELPTEETEIMIV QAKGRGRGRKETGTAKPGVSTVPNTTQASTPPQTQTP QPNPPPVQATPHFPFAVTPDLIVQTPVMTVVPQPLQ TPPPVPPQPPPPAPAPQPVQSHPIIAATPQPVKTK KGVKRKADTTTPTTIDPIHEPPSLPPEPKTTKLGQRR ESSRPVKPPKDVPSQQHPAPEKSSKVSEQLKCCSG ILKEMFAKKHAAYAWPFYKPDVEALGLHDYCDI IKH PMDMSTIKSKLEAREYRDAQEFGADVRLMFSNCYKYN PPDHEVVAMARKLQDVFMRFKMPDEPEEPVAVSS PAVPPPTKVVPSSSDSSSDSSSDSSSDSDSDSDSEER AORLAELQEQLKAVHEQLAALSQPPQNNPKKKKEDKK

630
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EKKKEKHKRKEEVEENKKS KAKEPPPKTKKNSSNS NVSKKEPAPMKS KPPPTYEESEEDKCKPMSYEEKRQL SLDINKLPGEKLGRVVHIIQSREPSLKNSNPDEIEID FETLKPSTLRELERVYVTSCLRKKRKPQ/ASEKVDVIA GSSKMKGFSSSESESSSSSSSDSEDSETGPA
2160	A	108	440	MQATSNLLNLLLLSLFAGLNPSKTHINPKEGWQVYSS AQDPDGRGICTVVAPEQNLC SRDAKSRQLRQLLEKVQ NMSQSIEVLNLR TQ RDFQYVLKMETQMKGLKAKFRQI
2161	A	18	467	REELGKDLFDCTLYVLLKYDDFNADKHLALEEFYRAF QVIQLSLPEDQKLSITAATVGQSAVLSCAIQGTLRPP IIWKRNNIILNNLDLEDINDFGDDGSLYITKVTTHV GNYTCYADGYEQVYQTHIFQVNVPPVIRVYPESQARR AG
2162	A	79	415	MFYQMIWTNGPAKL PASSTKHDLYLCNSFTGPSNIIW NLGSRYIFTVIKHGLGFFLNTILAVLNIAGRNLKCYK FC* TGWKLGWSIGPNHLIKHLQTVQONTIYIRRPSKG VAQVRTRGS
2163	A	59	447	ITVDRNTETRTSSFSIISVPASST*GSPSRVIYAKLG GEILDYRDLAALPKSKAIYDIDRPDMISYSPYISHSA GDRQSYGES PQLLSPTPTEGDQDDRSYKQCR TSSPSS TGLVSLGRYTPTS RAPQH
2164	A	3	493	DPRVRFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVI LYILLCGFPFPRSPERDQDELFNIIQLGHFEFLPPYW DNISDAAKDLVSRLLVVDPKKRYTAHQVLQHPWIETA G/EDQYSETTEAGVPQQRGSLPEPAQEGCGAGIIVTT LGICPAPSSAQGQRKG
2165	A	3	493	DPRVRFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVI LYILLCGFPFPRSPERDQDELFNIIQLGHFEFLPPYW DNISDAAKDLVSRLLVVDPKKRYTAHQVLQHPWIETA G/EDQYSETTEAGVPQQRGSLPEPAQEGCGAGIIVTT LGICPAPSSAQGQRKG
2166	A	1334	470	SAAQLSLCSRLQLTLYQYTTCPFCSGVRAFLDFHALP YQVVEVNPERRAEIKFSSYRKVPILVAQEGESSQQLN DSSVIIISALKTYLVSGQPLEEIIITYYPAMKAVNEQGK EVTEFGNKYWLMLNEKEAQQVYGGKEARTEEMKWRQW ADDWLVLH LISP NVYRTPT EALASFDYIVREGKFGAVE GAVAKYMGAAAMYLI SKRLKSRHRLQDNVREDLYEAA DKWVA AVGKDRPFMG GQKPNLADLAVYGVL RVMEGLD AFDDLMQH THIQPWYLRVERAITEASPAH
2167	A	996	214	GRIRMQRQSTTGGRGIMEGPRGWLVL CVLAISLASMV TEDLCRAPDGKKGEAGRPGRGRPGLKGEQGEFGAPG IRTGIQGLKGDQGEPPSGNPGKVGYPGPSGPLGARG IPGIKGTGKSPGNIKDQPRPAFSAIRRNPPMGNNVVI FDTVITNQEEPYQNHSGRFVCTVPGY YFTFQVLSQW EICLSIVSSSRGQVRRSLGFCDTTNKGLFQVVS GGMV LQLQQGDQVWVEKDPKKGHIYQGSEADSVFSGFLIFP SA
2168	A	3	420	LRRFSTDCSSDQQDR LNGTAPSGFNRS*PVPLPHPI L EVC PGQ*EPQSAISLTAFQVQAGASRASPGPPAPSSS KPGRKAKVASPCPDRPAPPPT*PRPAAAPGSESSPRP

631
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PRPRTGRRQQRAHARRAAARTAPWRPSC
2169	A	2744	496	ENEEQDSQNEGSTDEKSSPASSQEGSPSGDQQFSPKS NTEKSKGELMFDDSSDSSPEKQERNLNWTPAEVPQLA AAKRRLPQGKEPGLINLCANVPPVPGNILPPEVRGNL MAAGQNLQSSERSEMIATWSPAVRTLNRNITNNADIQQ MNRPSNVAHILQTLTSAPTKNLEQQVNHSQQGHTNANA VLFSQVKVTPETHMLQQQQQQAQQQQQHPVLHLQPQQ IMQLQQQQQQQISQQPYPPQQPPHPFSQQQQQQQQPP SPQQHQLFGHDPAVEIPEEGFLLGCVFAIADYPEQMS DKQLLATWKRI IQAHGGTVDP\PSRVDARTFSVRVKS AAR/IAQAIRERKRCVTAHWLNTVLKKKKMVPPhRAL HFPVAFPPGGKPCSQHIISVTGFVDSRDDDLKLMAYL AGAKYTGylCRSNTVLICKEPTGLKYEKAKEWRIPC NAQWLGDILLGNFEALRQIQYSRYTAFSLQDPFAPTQ HVLNLLDAWRVPLKVSaelLMSIRLPpKLKQNEVAN VQP\SSKRARIED\VPPPTKKLTP\ELTPF\VLFTGF EPVQVQQYI\KKLYILGGEVAESAQKCTHLIASKVTR TVKFLA\AISVVKHIVTPEWLEECFRCQKFIDEQNYI LRDAEAEVLFSFSLEESLKRAHVSPLFKAKYFYITPG \ICPSLSTMKAIVECAGGKVLsk\QPSFRKLMGAQAG TSSLFGK*F*LSC\ENDLHFIR\EYFARG\IDVHNAE F\VLTEVLTQTLdYESYKV
2170	A	2744	496	ENEEQDSQNEGSTDEKSSPASSQEGSPSGDQQFSPKS NTEKSKGELMFDDSSDSSPEKQERNLNWTPAEVPQLA AAKRRLPQGKEPGLINLCANVPPVPGNILPPEVRGNL MAAGQNLQSSERSEMIATWSPAVRTLNRNITNNADIQQ MNRPSNVAHILQTLTSAPTKNLEQQVNHSQQGHTNANA VLFSQVKVTPETHMLQQQQQQAQQQQQHPVLHLQPQQ IMQLQQQQQQQISQQPYPPQQPPHPFSQQQQQQQQPP SPQQHQLFGHDPAVEIPEEGFLLGCVFAIADYPEQMS DKQLLATWKRI IQAHGGTVDP\PSRVDARTFSVRVKS AAR/IAQAIRERKRCVTAHWLNTVLKKKKMVPPhRAL HFPVAFPPGGKPCSQHIISVTGFVDSRDDDLKLMAYL AGAKYTGylCRSNTVLICKEPTGLKYEKAKEWRIPC NAQWLGDILLGNFEALRQIQYSRYTAFSLQDPFAPTQ HVLNLLDAWRVPLKVSaelLMSIRLPpKLKQNEVAN VQP\SSKRARIED\VPPPTKKLTP\ELTPF\VLFTGF EPVQVQQYI\KKLYILGGEVAESAQKCTHLIASKVTR TVKFLA\AISVVKHIVTPEWLEECFRCQKFIDEQNYI LRDAEAEVLFSFSLEESLKRAHVSPLFKAKYFYITPG \ICPSLSTMKAIVECAGGKVLsk\QPSFRKLMGAQAG TSSLFGK*F*LSC\ENDLHFIR\EYFARG\IDVHNAE F\VLTEVLTQTLdYESYKV
2171	A	3	581	GRRLRSEPRPARPPIARAWPPAPGADGRARRTRVPAP CLPRAPCYGVRPRAWRPRLRGLVRWLLSGGPQP RRPRATERPSAGTGAAPRTEPRGRCRCGCRGRG*GP RAWGLALCSPHSCSGAAWGPTTGSQRSWPAVARSWQG DSSRCPALRTTTVTAGSKAALPESAAEVSPMSSSPGR KRSGFAA
2172	A	70	993	SEQKIQEQGYVWITVFSALPTTVSALHPRVLKPLSSL

632
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				IHLQANSNPWECNCKLLGLRDWLASSAITLNIYWQNP PSMRGRALRYINITNCVTSSINVSRAWAVVKSPIHH KTTALMMAWHKVTTNGSPLENTETENITFWERIPTSP AGRFFQENAFGNPLETTAVLPVQIQLTTSVTLNLEKN SALPNDAAASMSGKTSLICTEVEKLNEAFDILLAFFI LACVLIIFLIYKVQFKQKLKASENSRENRLYYSFY QSARYNVTASICNTSPNSLESPGLEQIRLHKQIVPEN EAQVILFEHSAL
2173	A	2	722	AVRLNISYPPQNLTMVTFQGDGTASTTLRNGSALSVL EGQSLHLVCAVDSNPPARLSWTWGSLLTSPSQSSNLG VLELPRVHVKDEGEFTCRAQNPLGSQHISLSLSLQNE YTGKMRPISGVMLGAFGGAGATALVFLSFCIIFVVVR SCRKKSARPAVGVDGTMEDANAVRGASQGPLIESP ADDSPPHHAPPALATPSPEEGEIQYASLSFHKARPQY PQEQEAIGYEYSEINIPK
2174	A	2043	1232	SHIQHHGRGAQAPVKMVSWMISRAVVLVFGMLYPAYY SYKAVKTKNVKEYVRWMMYWIWFALYTVIETVADQTV AWFPLYELKIAFVIWLLSPYTKGASLIYRKFLHPLL SSKEREIDDIYVQAKERGYETMVNFGRQGLNLAATAA VTAAVKSQGAITERLRSFSMHDLTITQGDPEVGQRPY QPLPEAKKKSKPAPSESAGYGIPLKDGDEKTDDEEAEG PYSDNEMLTHKGPRRSQSMKSVKTTKGRKEVRYGSLK YKVKKRPQVYF
2175	A	1	790	RGYNPNVNAGIINSFATAAFRFGHTLINPILYRLNAT LGEISEGHLPPFKALFSPSRIIKEGGIDPVLRLGLFGV AAKWRAPSYLLSPELTQRLFSAAYSAAVDSAATIIQR GRDHGIPPYVDFRVFCNLTSVKNFEDLQNEIKDSEIR QKLRLKLYGSPGDIDLWPALMVEDLIPGTRVGPTLMC/ ML/STQFQRLRDGDRFWYENPGVFTPAQLTQLKQASL SRVLCNNGDSIQQVQADVFRKRQEYPQDYLNCKRES PNVDPKAC
2176	A	1	790	RGYNPNVNAGIINSFATAAFRFGHTLINPILYRLNAT LGEISEGHLPPFKALFSPSRIIKEGGIDPVLRLGLFGV AAKWRAPSYLLSPELTQRLFSAAYSAAVDSAATIIQR GRDHGIPPYVDFRVFCNLTSVKNFEDLQNEIKDSEIR QKLRLKLYGSPGDIDLWPALMVEDLIPGTRVGPTLMC/ ML/STQFQRLRDGDRFWYENPGVFTPAQLTQLKQASL SRVLCNNGDSIQQVQADVFRKRQEYPQDYLNCKRES PNVDPKAC
2177	A	1	790	RGYNPNVNAGIINSFATAAFRFGHTLINPILYRLNAT LGEISEGHLPPFKALFSPSRIIKEGGIDPVLRLGLFGV AAKWRAPSYLLSPELTQRLFSAAYSAAVDSAATIIQR GRDHGIPPYVDFRVFCNLTSVKNFEDLQNEIKDSEIR QKLRLKLYGSPGDIDLWPALMVEDLIPGTRVGPTLMC/ ML/STQFQRLRDGDRFWYENPGVFTPAQLTQLKQASL SRVLCNNGDSIQQVQADVFRKRQEYPQDYLNCKRES PNVDPKAC
2178	A	501	187	AGVKWYEHGLWQPPPPGLKRSSHLSLPSS*DHRHEY CPANF*KIFF\VETRSHYVAQTSLEFLDSSNPPTSAS QNAGI*GMSHCAQPMQTFSLVKIGTNFLIF

633
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2179	A	4312	2359	AEKKMLPVDGEERKSEGS DTEGDR TSPCAVSS LIVSN RYPRGGPYII \ATLKDLEVGGSGRRCS DPAGQPSNLL PQRGLGAPLPAETAHTQPSPNDRSLYLSPKSSSSASSS LHARQSPCQEQA AVLNSRSIKISRLNDTIKSLKQKK QVEHQLEEEKKANNEKQKAERELEGQIQRLNIEKKKL NTDLYHMKHSLRYFEEESKDLA GR LQRSSQRIGELEW SLCAVAATQKKKPDGFSSRSKALLKRQLEQSIREQIL LKGHVTQLKESLKEVQLERDQYAEQIKGERAQWQORM RKMSQEVCTLKEEKKH DTHRVEELERSLSRLKNQMAE PLPPDAPAVSSEVELQDLRKE LERVAGELQAQVENNQ CISLLNRGQK \ERLREQEERLQEQQERLREREKRLQQ LAEPQSDLEELKHENKSALQLEQQVKELQEKLGQVME TLTSAEKEPEAAVPASGTGGESSGLMDLLEEKADLRE HVEKLELGF IQYRRERCHQNVHRL LTEPGDSAKDASP GGGHHQAGPGQGGEEGAAGAAGDGV AACGSYSEGHG KFLAAQNPAAEPSPGAPAPQELGAADKHGDLCEASL TNSVEPAQGEAREGSSQDNPTA \QPIVQLLGEMQDHQ EHPGLGSNCCVPCFCWAWLP RRRR
2180	A	2	1273	GGALQCGDPLARSPAVPAPRVPAQPPPGLGRRASRKE AATLAMASPPACPSEEDESLKGCELYVQLHGIQQVLK DCIVHLCISKPERPMKFLREHF EKELEKEENRQILARQ KSNSQSDSHDEEVSPTPPNPVVKARRRRGGVSAEVT EEDAVSYVRKVI PKDYKTMTALAKAISKNVLF AHLDD NERSDIFDAMFPVTHIAGETVIQQGNEGDNFYVVDQG EVDVYVNGEWTNI SEGGSFGELALIYGT PRAATVKA KTDLKLWGIDRDSYRRILMGSTLRKRKMYEEFLSKVS ILESLEKWERLTVADALEPVQFEDGEKIVVQGE PGDD FYIITEGTASVLQRRSPNEEYVEVGR LGPSDYFGEIA LLLNRPRAATVVARGPLKCVKLD RPRFERVLGPCSEI LKRNIQRYNSFISLTV
2181	A	1	303	PTRPLERGPSGLGMGLIDGMH THLGAPGLYIQTLLPG SPAAADGRLSLGDRILEVNGSSLLGLGYLRAVDLIRH GGKKMRFLVAKSDVETAKKIHFRTPPL
2182	A	2227	332	MGKYTVRVATGDLLLAGSPNLVQLWL VGEHGEADLGK QLPPVWGKEAEFEIDVPLHLGRLLMVKLRKHNVLLSL DWFCWKISVQGPQTQGAFFPCYRWVQGHG IICLPEG T/RWGSWKDGLILPIAGNRQPD LPRDERFLEDKD LDF NVSLAKGLKDLAIKGTLD F INCVKRLED FKKIFPHGK TVLAERVYDSWKND AFFGYQFLNGANPMLLRCSRLP ACLVLP PGMEDLKTQLEKELQAGSLFEVDFSLLDGVK PNVII FKQQCVAAPLVVLKLQPDGGLLPMVIQLQPP* HGCPPPLLFLPSHPPMAWLLAKTWVRSSDFQLQQLQS HLLRGHLIAEVI AVATMRSLPSLHPIYKLLI PHFRYT MAINTLAQSSLVSEWGI FDLVVSTGSGSHVDILQAM ACLTYSLSLCPDDLADRGLLDVKSSFY G*DAIRLWGI ISRE* \YVEGMVGLFYNSDQAMKDDLELQAWCREMTE TGLQRAQDQGFLISLESRAQLCHFVTMCIFTCTGQHA SNHLGQLDWYSWIPNGPCTMQKPPPI SKDVTEKDIVD LLPNLHQARMQKTFTKFLGRRQPMHEEKYFSGPEPQ AVLRQFQEELASMDKEIEVRNAV LNLPC EYL*PSMVE

634
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				\NSVTI
2183	A	2227	332	MGKYTVRVATGDLLLAGSPNLVQLWLVGEGEADLGK QLPPVWGKEAEFEIDVPLHLGRLLMVKLRKHNVL LSL DWFCWKISVQGPQTQGAFFPCYRWVQGHGII CLPEG T/RWGSWKDGLILPIAGNRQPDLP RDERFLEDKD LDF NVSLAKGLKDLAIKGTLD F INCVKRLED FKKI FPHGK TVLAERVYDSWKND A FFGYQFLNGANPMLLRCS S RLP ACLVLP PGMEDLKTQLEKELQAGSLFEVDFSLLDGVK PNVII FKQQCVAAPLVVLKLQPDGGLLPMVIQLQPP* HGCPPPLLFLPSHPPMAWLLAKTWVRSSDFQLQQ LQS HLLRGHLIAEVI AVATMRS LPSLHPIYKLLI PHFRYT MAINTLAQSSLVSEWGI FDLVVSTGSGSHVDILQ RAM ACLT YHSLCPPDDLADRGLLDVKSSFYG*DAIRLWGI ISRE*\YVEGMVGLFYNSDQAMKDDLELQAWCREMTE TGLQRAQDQGFLISLESRAQLCHFVTMCIFTCTGQHA SNHLGQLDWYSWIPNGPCTMQKPPPI SKDVTEKD IVD LLPNLHQARMQKTFTKFLGRRQPMHEEKYFSGPEPQ AVLRQFQEELASMDKEIEVRNAVLNLPCEYL*PSMVE \NSVTI
2184	A	2227	332	MGKYTVRVATGDLLLAGSPNLVQLWLVGEGEADLGK QLPPVWGKEAEFEIDVPLHLGRLLMVKLRKHNVL LSL DWFCWKISVQGPQTQGAFFPCYRWVQGHGII CLPEG T/RWGSWKDGLILPIAGNRQPDLP RDERFLEDKD LDF NVSLAKGLKDLAIKGTLD F INCVKRLED FKKI FPHGK TVLAERVYDSWKND A FFGYQFLNGANPMLLRCS S RLP ACLVLP PGMEDLKTQLEKELQAGSLFEVDFSLLDGVK PNVII FKQQCVAAPLVVLKLQPDGGLLPMVIQLQPP* HGCPPPLLFLPSHPPMAWLLAKTWVRSSDFQLQQ LQS HLLRGHLIAEVI AVATMRS LPSLHPIYKLLI PHFRYT MAINTLAQSSLVSEWGI FDLVVSTGSGSHVDILQ RAM ACLT YHSLCPPDDLADRGLLDVKSSFYG*DAIRLWGI ISRE*\YVEGMVGLFYNSDQAMKDDLELQAWCREMTE TGLQRAQDQGFLISLESRAQLCHFVTMCIFTCTGQHA SNHLGQLDWYSWIPNGPCTMQKPPPI SKDVTEKD IVD LLPNLHQARMQKTFTKFLGRRQPMHEEKYFSGPEPQ AVLRQFQEELASMDKEIEVRNAVLNLPCEYL*PSMVE \NSVTI
2185	B	1	1110	MGLLICLGALDARPERAPSACGEVWRERRGREPGLPT VLAGQREFWVGVSAAHLSER PAGPTTPGSKGLSTQV SSCGGRTGSPSSASPLALRSISRWGLSHLPHGAGLRT CSPAMPKPPHSAVGSCATRASLISTAPRSRAPGPIDH PRAETCQRTVQELAGSSTCSPVQDPLGEASWAPEFEG SGPKRRANGRGAYGLRDTGVHSSGVAARSPAAAERWV QGFPKQNVHFVNDNTICYPCGNYVIFINIETKKKTVL QCSNGIVGVMATNIPCEVVAFS DRKLKPLIYVYSFPG LTRRTKLKADQERDPFLYLFQVAEFLTQGCLQISAFS PTSQRYQALLGQMWDLIRGHRFSVEKSVETSSSSCSA
2186	A	22	960	ARPGPDMAALYACTKCHQRFPFEALSQGGQLCKECRI AHPVVKCTYCRTEYQQESKTN TICKKCAQNVQLYGT KPCQYCNIIA AFIGNKCQRCTNSEKKYGPYPYSCEQCK

635
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QQCAFDRKDDRKKVDGKLLCWLCTLSYKRVLQKTKEQ RKHLSSSSSRAGHQEKEQYSRLSGGGHYNSQKTLSTSS IQNEIPKKKSKFESITTINGDSFSPDLALDSPGTDHFV IIAQLKEEVATLKKMLHQKDQMLEKEKKITELKADF QYQESQMRAMNQMEKTHKEVTEQLQAKNRELLKQAA ALSKSKKSEKSGAITSP
2187	A	612	812	RSGRTVVTGIGYSKALQSSNRNTKSLQNEFMMVYSF RALSFKESTWATFQHGGEATKSRSLSSTQ
2188	A	612	812	RSGRTVVTGIGYSKALQSSNRNTKSLQNEFMMVYSF RALSFKESTWATFQHGGEATKSRSLSSTQ
2189	A	612	812	RSGRTVVTGIGYSKALQSSNRNTKSLQNEFMMVYSF RALSFKESTWATFQHGGEATKSRSLSSTQ
2190	A	612	812	RSGRTVVTGIGYSKALQSSNRNTKSLQNEFMMVYSF RALSFKESTWATFQHGGEATKSRSLSSTQ
2191	A	612	812	RSGRTVVTGIGYSKALQSSNRNTKSLQNEFMMVYSF RALSFKESTWATFQHGGEATKSRSLSSTQ
2192	A	936	745	RRNSPGLCFLLP SLFHLRLLLWRLLLWHQVFFDVAIFV IGGICSVSGFVHSLEGLIEAYRTNAED
2193	A	122	643	MPSGCRCLHLVCLLCILGAPGQPVRADDCSSHCDLAH GCCAPDGSCRCDPGWEGLHCERCVRMPGCQHGTCHQP WQICHSWAGKFCDEHICTTQSPCQNGGQCMYDG GGEYHCVCLPGFHRDCERKAGPCEQAGSPCRNGGQC QDDQGFALNFTCRCLVGFVGARCDV*
2194	A	1	1406	NVVSRAAPPAPVEDLSKVS YEELLQWSKEELIRSLRA EAEKVSAMLDHSNLIREVNRRLQLHLGEIRGLKDINQ KLQEDNQELRDLCCFLDDDRQKGKRVSRWQRLGRYT AGVMHKEVALYLQKLKDLEVKQEEVVKENMEL/KELC VLLDEEKGAG\ SQAAAAPSTARPACANSQP/PTAPYV RDVGDGSSTSTGSTDSPDHHKHHASSGSP EHLQKPR SEGSP EHSKHSASPEHPQKPRACGTPDRPKALKGPS PEHHKPLCKGSPEQQRHPHPGSSPETLPKHVLSGSPE HFQKHSRSGSSPEHARHSGGSPEHLQKHALGGSLEHLP RARGTSPEHLKQHYGGSPDHKHGSGSGSGSGSGSR EGTLRRQAQEDGSPHHRNVYSGMNESTLSYVRQLEAR VRQLEENRMLPQASQNTGRPPTKNSSHMEKGWGSRA RRVLHWWQGCRGIGRCLATLTGSFRWSS
2195	A	1461	197	GVTHLFLFGKRKLRLNGIAEDLKGQADFFLLVSEAVV ATGSPRAWLTCLILPLPGIIFSVLPKAMSRPLLTFT PATDPSDLWKDGQQQPQPEKPESTLDGAAARAFYEAL IGDESSAPDSQRSQTEPARERKRKKRRIMKAPAAEAV AEGASGRHGQGRSLEAEDKMT HRI LRAAQEGDLPELR RLLEPHEAGGAGGNINARDAFWWTPLMCAARAGQGAA VSYLLGRGAAWVGVCELSGRDAAQLAEEAGFP EVARM VRESHGETRSPENRSPTPSLQYCENDTHFQDSNHRT STAHLLSLSQGPQPPNLPLGVPISSPGFKLLLRGWE PGMGLGPRGEGRANPIPTVLKRDQEGLYRSAPQPRV THFPWDTRAVAGRE\TPPRVATLSWREERRREE\KD RAWERDLRTYMNLEF
2196	A	10	768	SFAGAAARPSTPPASGRGAAPGRPGPSPMDLRAGDSW GMLACLCTVLWHLPAVPALNRTGDPGPGPSIQKTYDL

636
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TRYLEHQLRSLAGTYLNYLGPPFNEPDPNPPRLGAET LPRATVDLEVWRSINDKLRLTQNYEAYSHLLCYLRGL NRQAATAELRRSLAHFCTSLQGLLGSIAGVMAALGYP LPQPLPGTEPTWTTPGPAHSDFLQKMDDFWLLKELQTW LWRSKDFNRLKKKMQPPAAAVTLHLGAHGF
2197	A	1	1054	PPIARLQEFGTSSRRHMAAPSGVHLLVRRGSHRIFSSP LNHIYLHKQSSSQQRNFFFRQRDISHSIVLPAAVS SAHPVPKHIKKPDYVTTGIVPDWGDSEVKNEQIQG LHQACQLARHVLLLAGKSLKVDMTTEEIDALVHREII SHNAYPSPLGYGGFPKSVCTSVNNVLCHGIPDSRPLQ DGDIIINIDVTVYYNGYHGDTSFTLVGNVDECGKKLV EVARRCRDEAIAACRAGAPFSVIGNTISHITHQNGFQ VCPHFVGHGIGSYFHGHPEIWHHANDSDLPMEEGMAF TIEPIITEGSPEFKVLEDAWTVVSLD/TSKVSAQFEH TVLITSRGAQILTKLPHEA
2198	A	2319	957	SPGTPAAGRTSRTVQTPF*SRTPLALMIGSENWPGLQ /FPAKWAP*ANHLTFAGLTPNHSGTK\WAGISGTRLS LPGAGAAPEVPRRCRRHCPECLQPAGNAAPEQSGGC RLAFL*ARSTSSRARGLLGSEVRRPGVAGSQRAKLLT P*LPFLLGVSSPSPKSGSRTAAMHQPRLSPIQRRRK CSGEREASHYEPALSKAVRSVGGSPKSASGDAGRARS \SRAPNSESSNMAARLAIEREEKAGD*QAARRRRGPP PPFTSGI*SRLPEAGTMSA*QPTLEFGG/SLP*SKGN SSHKELEASPSVVGROPGAV\SGNCGMCPWGPEKTE GRCSRPTTAWCSLCSGCCPMTSLSIQNSCKRLL SSSLCSSSSRILQSSSTSSSFSSCSSTPSSSRLAWST SYSISSKGPSS*QLCTLPSASPFMSG*TYAGKTPTA SYGQMDFKCCLYSRD
2199	A	1	3349	MDQPEAPCSSTGPRLAVARELLLALEELSQEQLKRF RHKL RDVGPDGRSIPWGRLERADAVDLAEQLAQFYGP EPALVARKTLKRADARDVAAQLQERRLQRLGLSGT LLSVSEYKKKYREHVLQLHARVKERNARSVKITKFT KLLIAPESAAPEEALGPAAEPEPGRARRSDTHTFNRL FRDEEGRRPLTVVLQGPAGIGKTMAAKKILYDWAAG KLYQGQVDFAFFMPCGELLERPGTRSLADLILDQCPD RGAPVPQMLAQPRLLFILDGADELPAJGGPEAAPCT DPFEAASGARVLGGLLSKALLPTALLLVTTAAAPGR LQGRLCSPQCAEVRGFSKDKKKYFYKFFRDERRAER AYRFVKENETLFCFVFPVCWIVCTVLRQQLGRD LSRTSKTTTTSVYLLFITSVLSSAPVADGPRLOGLRN LCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLS KKELPGVLETEVTYQFIDQSFQEFALAALSYLLEDGGV PRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAER MRDIERHFHGMVSEKQALRWVQGGQGGCPGVAP VTEGAKGLEDTEPEEEEEEGEENYPLELLYCLYETQ EDAFVRQALCRFPALALQVRVFCRMDVAVLSYCVRC PAGQALRLISCLVAAQEKKKKSLGKRLQAR\LGGGS WLGTLAQPEVPFRPPCCDICPTPPDPRLQKAFAR VPLNIAPIQPLPRGLASVERMNVTVLAGAGPGDPKTH AMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPAL

637
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLP DPQRGLQYLVGMLRQSPALTTLDLSCQLPAPMVTYL CAVLQHQCGLQTLSSLPSDPTSSFSGRCREPGRR LGLESRWPRSAPEPSGRQRGEDPGGGGRGRGRREEAR EGTPGPRAPPTAAPGRSSGSRLELC SLRALRAGNARP PDATHAAAASGDRGEFGPRPRVHVPPPGPAQRPPPPP RDRPRLPATARALGAGTADLPGGAAAGRLLLP PGPGV EQRDTGSHAGARRPGGAAAQAQQLHGRRRGPHHVC CPLSAQ
2200	A	877	446	GIRCRFGTSEIRAHATAKATVAAFTASEGHAHPRVVE LPKTDEGLGFNIMGGKEQNSPIYISRVIPGGVADRHG GLKRGDQLLSVNGVSVEGEQHEKAVELLKAAQGSVKL VVRYPTRVLEEMEARFEKMRSARRRQQHQSYS
2201	A	48	474	SCLARPFRAQVSSSGFRAQNFPGVGSWAVAVGAGMAQ LEGYCFSAALSCTFLVSCLLFSAFSRALREP\YMDEI FHLPPAQRYCEGHFSLSQWDPMITTLPGLYLVSVGVV KPAIWI FGWSEHVVC SIGMLRFVNLLFSVGNF
2202	A	3140	1502	FRRLHSVPRGSALCAMDGI VPDIAVGTKRGSDELFT CVTNGPFFIMSSNSASAANGNDSKKFKGDSRSAGVPSR VIHIRKLPIDVTEGEVISLGLPFGKVTNLLMLKGKNQ AFIEMNTEEAANTMVNYTSTVTPVLRGQPIYIQFSNH KELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAA VDAGMAMAGQSPVLR I IVENLFYPVTL DVLHQI FSKF GTVLKIITFTKNNQFQALLQYADPVSAQHAKLSLDGQ NIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLF SGDSQPSLDQTMAAAFGLSVPNVHGALAPLAIP SAAA AAAAAGRIAI PGLAGAGNSVLLVSNLNPVRTPQSLF ILFGVYGDVQVRKILFNKKENALVQ MADGNQAQLAMS HLNGHKLHGKPIRITLSKHQNVQLPREGQEDQGLTKD YGNSPLHRFKKPGSKNFQNI FPPSATLHLSNI PPSVS EEDLKVLFS SNGGVVKGF KFFQKDRKMALIQMG SVEE AVQALIDLHNHDLGENHHLRVSF SKSTI
2203	A	2240	506	RRPPEGSGGGRRTRARMPLPWSLALPLLLSWVAGGF GNAASARHHGLLASARQPGVCHYGTKLACCYGWRRNS KGVCEATCEPGCKFGECVGPKNKCRCFPGYTGKTCSQD VNECGMKPRPCQHRCVNT HGSYKCFCLSGHMLMPDAT CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN GRDCLDIDECASGKVICPYNRRCVNTFGSYCKCHIG FELQYISGRYDCIDINECTMDSHTCSHHANCFNTQGS FKCKCKQGYKGNGLRCSAIPENSVKEVLRAPGTIKDR IKKLLAHKNSMKKKAKIKNVTPEPTRTPTPKVNLQPF NYEEIVSRGGNSHGG\KKGNEEKMKEGLEDEKREEKA LKD*HRRERPFRG\DVFFPKVNEAGEFGLIL\VQRKA LTSKLEHKADLNI SVDCSFNHG\ICDW\KQDR\EDDF DW\NPADR\DNAI\GFY\MAVPGLWQGHK\KDIGRLK LLLPDLQPQSNFCLLF DYRLAGDKVGKLRVFKNSNN ALAWKTTSEDEKWKTKGIQLYQGT DATKSI IFEAER GKGKTGEIAVDGVLLVSGLC PDSSL SVDD
2204	A	2240	506	RRPPEGSGGGRRTRARMPLPWSLALPLLLSWVAGGF GNAASARHHGLLASARQPGVCHYGTKLACCYGWRRNS

638
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				KGVCEATCEPGCKFGCECVGPNKCRCPGTYTGKTCSQD VNECGMKPRPCQHRVCVNTHGSKYCFCLSGHMLMPDAT CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN GRDCLDIDECASGKVICPYNRRCVNTFGSYCKCHIG FELQYISGRYDCIDINECTMDSHTCSHHANCFNTQGS FKCKCKQGYKGNLRCSAIPENSVKEVLRAPGTIKDR IKKLLAHKNSMKKKAKIKNVTPEPTRTPTPKVNLPF NYEEIVSRGGNSHGG\KKGNEEKMKEGLEDEKREEKA LKD*HRRERPFRG\DVFFPKVNEAGEFGLIL\VQRKA LTSKLEHKADLNISVDCSFNHG\ICDW\KQDR\EDDF DW\NPADR\DNAI\GFY\MAVPGLWQGHK\KDIGRLK LLLPDLQPQSNFCLLDYRLAGDKVGKLRVFKNSNN ALAWEKTTSEDEKWKTKIQLYQGTDATKSIIFEAE GKGTGEIAVDGVLLVSGLCPDLSLLSVDD
2205	A	2814	346	VKKTKSIFNSAMQEMEYVENIRRKFGVFNYSFRTTP YTPNSQYQMLLDPTNPSAGTAKIDKQEKVKLNFDMTA SPKILMSKPVLSGGTGRRISLSDMPRSPMSTNSSVHT GSDVEQDAEKKATSSHFSASEESMDFLDKSTASPAST KTGQAGSLSGSPKPFSPQLSAPITTKTDKTSTTGSIL NLNLDRSKAEMDLKELSES VQQQSTPVPLISPKRQIR SRFQLNLDKTIESCKAQLGINEISEDVYTAVEHSDSE DSEKSDSSDSEYISDDEQKS*GTSQEDTEDKEGCQMD KEPSAVKKKPKPTNPVEIKEELKSTSPASEKADPGAV KDKASPEPEKDFSGKAKPSPHPIKDKLKGKDETDSP VHLGLDSDSE\NELVIDLGEDHSGREGRKNKKEPKEP SPKQDVVGKTPPSTTVGSHSPPETPVLTRSSAQTSAA GATATTSTSSSTVTVTAPAPAATGSPVKKQRPLLPKE\ TAPAVQRSCGTSSSTVQQKEITQSPSTSTITLVTSTQS SPLVTSSGSMSTLVSSVNGDLPIGTASADVAADIAY TSKL\MDAIKGTM\TEIYNDLSKN\TTWKAQLAEDSQ GLRIEIEKLQWLHQEL\SEMKNLELTMAEMRQSW QERDRLIAEVKKQLELEKQQAQVDETKKKQWCANFKKE AIFYCCWNTSYCDYPCQ\QAHWPEH\MKSCQTQSATA P\Q\QEADAE\VNTETLNKSSQGSSSSTQSAPSETASA\ SKEKETSAEKSKESGSTLDLGSRETPSSILLGSNQ SDHSR\SNKSSWSSSDEKRG\TRSDHN\TPSTQHGR SLLPGKESRAGTPFLGTSK
2206	A	2814	346	VKKTKSIFNSAMQEMEYVENIRRKFGVFNYSFRTTP YTPNSQYQMLLDPTNPSAGTAKIDKQEKVKLNFDMTA SPKILMSKPVLSGGTGRRISLSDMPRSPMSTNSSVHT GSDVEQDAEKKATSSHFSASEESMDFLDKSTASPAST KTGQAGSLSGSPKPFSPQLSAPITTKTDKTSTTGSIL NLNLDRSKAEMDLKELSES VQQQSTPVPLISPKRQIR SRFQLNLDKTIESCKAQLGINEISEDVYTAVEHSDSE DSEKSDSSDSEYISDDEQKS*GTSQEDTEDKEGCQMD KEPSAVKKKPKPTNPVEIKEELKSTSPASEKADPGAV KDKASPEPEKDFSGKAKPSPHPIKDKLKGKDETDSP VHLGLDSDSE\NELVIDLGEDHSGREGRKNKKEPKEP SPKQDVVGKTPPSTTVGSHSPPETPVLTRSSAQTSAA GATATTSTSSSTVTVTAPAPAATGSPVKKQRPLLPKE\ TAPAVQRSCGTSSSTVQQKEITQSPSTSTITLVTSTQS SPLVTSSGSMSTLVSSVNGDLPIGTASADVAADIAY TSKL\MDAIKGTM\TEIYNDLSKN\TTWKAQLAEDSQ GLRIEIEKLQWLHQEL\SEMKNLELTMAEMRQSW QERDRLIAEVKKQLELEKQQAQVDETKKKQWCANFKKE AIFYCCWNTSYCDYPCQ\QAHWPEH\MKSCQTQSATA P\Q\QEADAE\VNTETLNKSSQGSSSSTQSAPSETASA\ SKEKETSAEKSKESGSTLDLGSRETPSSILLGSNQ SDHSR\SNKSSWSSSDEKRG\TRSDHN\TPSTQHGR SLLPGKESRAGTPFLGTSK

639
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TAPAVQRSCGTSSTVQQKEITQSPSTSTITLVTSTQS SPLVTSSGSMSTLVSSVNGDLPIGTASADVAADIAKY TSKL\MDAIKGTM\TEIYNDLSKN\TTWKAQLAEDSQ GLRIEIEKLQWLHQEL\SEMKNLELTMAEMRQSWE QERDRLIAEVKKQLELEKQQAVDETKKKQWCANFKKE AIFYCCWNTSYCDYPCQ\QAHWPEH\MKSCTQSATAP Q\QEADAE\VNTETLNKSSQGSSSSTQSAPSETASA\ SKEKETSAEKSKESGSTLDLGSRETPSSILLGSNQG SDHSR\SNKSSWSSSDEKRGs\TRSDHN/TPSTQHGR SLLPGKESRAGTPFLGTSK
2207	A	2814	346	VKKTCSI FNSAMQEMEYVENIRRKFGVFNYS PFRT P YTPNSQYQMLLDPTNPSAGTAKIDKQEKVKLNFDMTA SPKILMSKPVLSGGTGRRISLSDMPRSPMSTNSSVHT GSDVEQDAEKKATSSHFSASEESMDFLDKSTASPAST KTGQAGSLSGSPKPFSPQLSAPITTKTDKTSTTGSI L NLNLDRSKAEMDLKELSES VQQQSTPVPLISP KRQIR SRFQLNLDKTIESCKAQLGINEISEDVYTAVEHSDSE DSEKSDSSDSEYISDDEQKS*GTSQEDTEDKEGCQMD KEPSAVKKKPKPTNPVEIKEELKSTSPASEKADPGAV KDKASPEPEKDFSGKAKPSPHPIKDKLKGKDETDSP T VHLGLDSDSE\NELVIDLGEDHSGREGRKNKKEPKEP SPKQDVVGKTPPSTTVGSHSPPETPVLTRSSAQTSAA GATATTSTSSSTVTVTAPAPAATGSPVKKQRPLLPKE\ TAPAVQRSCGTSSTVQQKEITQSPSTSTITLVTSTQS SPLVTSSGSMSTLVSSVNGDLPIGTASADVAADIAKY TSKL\MDAIKGTM\TEIYNDLSKN\TTWKAQLAEDSQ GLRIEIEKLQWLHQEL\SEMKNLELTMAEMRQSWE QERDRLIAEVKKQLELEKQQAVDETKKKQWCANFKKE AIFYCCWNTSYCDYPCQ\QAHWPEH\MKSCTQSATAP Q\QEADAE\VNTETLNKSSQGSSSSTQSAPSETASA\ SKEKETSAEKSKESGSTLDLGSRETPSSILLGSNQG SDHSR\SNKSSWSSSDEKRGs\TRSDHN/TPSTQHGR SLLPGKESRAGTPFLGTSK
2208	A	2814	346	VKKTCSI FNSAMQEMEYVENIRRKFGVFNYS PFRT P YTPNSQYQMLLDPTNPSAGTAKIDKQEKVKLNFDMTA SPKILMSKPVLSGGTGRRISLSDMPRSPMSTNSSVHT GSDVEQDAEKKATSSHFSASEESMDFLDKSTASPAST KTGQAGSLSGSPKPFSPQLSAPITTKTDKTSTTGSI L NLNLDRSKAEMDLKELSES VQQQSTPVPLISP KRQIR SRFQLNLDKTIESCKAQLGINEISEDVYTAVEHSDSE DSEKSDSSDSEYISDDEQKS*GTSQEDTEDKEGCQMD KEPSAVKKKPKPTNPVEIKEELKSTSPASEKADPGAV KDKASPEPEKDFSGKAKPSPHPIKDKLKGKDETDSP T VHLGLDSDSE\NELVIDLGEDHSGREGRKNKKEPKEP SPKQDVVGKTPPSTTVGSHSPPETPVLTRSSAQTSAA GATATTSTSSSTVTVTAPAPAATGSPVKKQRPLLPKE\ TAPAVQRSCGTSSTVQQKEITQSPSTSTITLVTSTQS SPLVTSSGSMSTLVSSVNGDLPIGTASADVAADIAKY TSKL\MDAIKGTM\TEIYNDLSKN\TTWKAQLAEDSQ GLRIEIEKLQWLHQEL\SEMKNLELTMAEMRQSWE

640
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QERDRLIAEVKKQLELEKQQAVDETKKKQWCANFKKE AIFYCCWNTSYCDYPCQ\QAHWPEH\MKSCTQSATAP Q\QEADAE\VNTETLNKSSQGSSSSTQSAPSETASA\ SKEKETSAEKSKESGSTLDLSGSRETPSSILLGSNQG SDHSR\SNKSSWSSSDEKRG\TRSDHN/TPSTQHGR SLLPGKESRAGTPFLGTSK
2209	A	1	575	GGIPHYLRGVNNARQPWHNADVRLRYGLRPGNATEEG LASLHSVLFKQPFLLWRAALLYTIHRAARMSFRQLF QDLERYVQDADVRWEYCVRAKRGQTDTSPLGCFSKDQ VYLDGIVRILRHRQTIDFPLLTSLGKVSIEDVDHLRP HGVLDNTRVPHFMQDLARYRQQLEHIMATNRLDEAEL GRLLPD
2210	A	3	1795	LGLSGTLLSVSEYKKKYREHVLQLHARVKERNARSV KITKRFTKLLIAPESAAPPEEALGPAAEEPEPGRARRSD THTFNRLFRRDEEGRRPLTVVLQGPAGIGKTMAAKKI LYDWAAGKLYQGQVDFAFFMPCGELLERPGTRSLADL ILDQCPDRGAPVPQMLAQPQRLLFILDGADELPAALGG PEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT RAAAPGRLQGRLCSPQCAEVRGFSKDKKKYFYKFFR DERRAERAYRFVKENETLFCFVPFVCWIVCTVLRQ QLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPR LQGDRLNLCRLAREGVIGRRRAQFAEKELEQLELRGSK VQTLFLSKKELPGVLETEVTYQFIDQSFQEFALAALSY LLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLF GLLSAERMRIERHFGCMVSERVKQEALRWVQGGQGG CPGVAPPEVTEGAKGLEDTEEPPEEEEEEGEPPNYPLELL YCLYETQEDAFVRQALCRFPPELALQVRVFCRMDVAVL SYCVRCCPAGQALRLISCLVAAQEKKKKSLGKRLQA SLGGG
2211	A	2	1177	GFVEAGEECYCVS\GQECRDLCCFAHNCSLRPGAQCA HGDCCVRCLLKPGALCRQAMGDCDLPEFCTGTSSH PPDVYLLDGSPCARGSGYCWGACPTLEQQCQQLWGP GSHPAPEACFQVVSAGDAHGNCQDSEGHFLPCAGR DALCGKLQCQGGKPSLLAPHMVPVDSTVHLDGQEVTC RGALALPSAQLDLLGLGLVEPGTQCGPRMVCQSRRCR KNAFQELQRCLTACHSHGVCNSNHNCHCAPGWAPPFC DKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLPLLPG AGLAWCCYRLPGAHLQRCSWGCRDPACSGPKDGP DHPLGGVHPMELGPTATGQPWPLDPENSHEPSSHPEK PLPAVSPDPQADQVQMPRSCLW
2212	A	1073	480	XXPDALSTVAEXPGRPTRPPTRTAAPWPRPGCSSASA PPTPASAPWPASPSSSSGRWSTDGRGPRPWEGSQGCW HCGSW*RT*CTCKIIGGPGSRGCAASSSWASSSRPSP SLPSAPSSCWPSPGIRASQTPPATTSPPASGASFPSSG PSCSASMPTATGLTLLTSASSAISDPGGSVYA*SGMV HQSGKEPSTVYTS
2213	A	1	2454	MALQNALYTGDLARLQELFPPHSTADLLLESRAAEPR WSSHQRACPIAYTLAQEHSHVEPRIAPAGCVARLVEK PSRGSEEHKSGPGPIVTRTASGPALAFWQAVLAGDV GCVSRILADSSTGLAPDSVFDTSDFRWRDFRNFIRA

641
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LRLWSLTYYYEELTTPLHVAASRGHTEVLRLLLLRRRAR PDSAPGGRTALHEACAAGHTACVHVLLVAGADPNIA QDGKRPLHLRCRPGTLECAELLLRFGARVDGRSEEEE ETPLHVAARLGHVELADLLLRGACPDARNAEGWTP LAACDVRCQSITDAEATTARCLQLCSLLLSAGADADA ADQDKQRPLHLACRRGHAAVVELLLSCGVSANTMDYG GHTPLHCALQGPAALAQSPHEVVRALLNHGAVRVWP GALPKVLERWSTCPRTIEVLMNTYSVVQLPEEAVGLV TPETLQKHQRFYSSLFALVRQPRSLQHLSRCALRSHL EGSLPQALPRLPLPPRLRLRYLQLDFEGVSPGICEQSQ LLGVQGCVEGKRRVGEGPSQNRVPPEPPEASESKPLL PDVHGLLRGPESRCPSLQRARLCTNSGQVALAAGGPA PQAGVDAI PNAEKRTDSGSRPFQGLLRSGTAHGGKD CPPGPHQVRLAGSRSAHRRKRQLCAAATRGHPRPGP TLPTMRGLSLANEWIGASFAGRLTNTFCAGLGQAVPS MVALTTALPSFAEPDPAFYGPQELAAAAAAAAATAAR NNPEPGRRPEGGLEADELLPAREKVAEP PPPPPPHF SETFPSLPGVDKLQGWDFRGHQDGGMLKQLSIQQWRA RSGF
2214	A	757	208	NVFI EPRIQGF MKTSAHPGQKHPDFSMGLLFPLLAAL EVCSCGSSGSLGYNL PQNH\GLLGRNTLVLLGQMRRI SPFLCLKDRSDFRFPQEKVEVSQLOKA\QAMSFYLDV LQQVFNF SHKALL\CCMEHDLPGTPHFTSSAAGTPG DLLGAGDGRRRSWGQWVIEGSTLALRRYFQESISTLE
2215	A	43	1004	QLWGFAAGSDSRPAMGCDGGTIPKRHEL VKGP KKEK VDKDAELVAQWNYCTLSQEILRRPIVACELGRLYNKD AVIEFLDKSAEKALGKAASHIKSIKNVTELKLSNPN AWEGDKGNTKGDKHDDLQRRARFICPVVGLEMNGRHRF CFLRCCGCVFSEALKEIKAEVCHTCGA AFQEDDVIV LNGTKEDVDVLKTRMEERRLRAKLEKKT KPKAESV SKPDVSEEAPGPSKVKTGKPEEASLDSREKKTNLAPK STAMNESSSGKAGKPPCGATKRSIADSESEAYKSLF TTHSSAKRSKEESAHWVTHTSYCF
2216	A	1323	840	FCPLGKPV MGPIFLDCRPFFLFPKPNQGTGTPLHNKV PYFFQ*GPFGLWNHRTLFFF LRWSFALLAQAGVQWR DLGSLQPLPPGFK*FSCLSLPSIWDYRRLPPCPANFA FLVETGFLHVGQVGL*LLTSGDPSASASQSSGITG\V SHHTWP*LSFLLWI
2217	A	17	348	ARAAARAGFSSYLKSLPDVRKKS LPLPEKPHKEENSE IVVWREFDKQVFLN*SPRRQSKLYTVDL ESGLHYLL RVELAAHKS LAGAELKTLKDFVTVLAKLFPGRPPVK
2218	A	1	1206	MALSSWPVVLRLNMADFVFSFLCLGIGTSIVLGILFY LLQAHRYLQEGMTYQLALSFYLTWASVFLFLMTGMGE DEESALQTL LDPRSSYLLVSLEILPTNPSPLSPCAVS EDESEMRLSLLRRQSQATGRLEPTFKHDSTLLALQG ALGLYDGHTPPYAACLGF EFRKHLGNPAKDGGNVTVS LFYRNDSAHLPLPLSLPGCPAPCPLGRFYQLTAPARP PAHGVSCHGPYEAVIPPGPGAIIPSTGPAVGMQRERS EVGSGVPARTVYASEQHAYMWHSALIPDSGLRGKPTL SSRKPPQTSCGPEFANVLSLALCGALVVCKARAMDQA

642
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				RPRQLIGIDALRDP RASSRTRAGGLGMIRRQEEEPAA RTVLARCDSSPSECPSHARAPYDTGPLFNAKG
2219	A	1	1594	NGGGSLNNYSSPIPSTPAPSR RDPQFSVPPTANTPTP VCKRSMRWSNLFTSEKGS DPKERKAPENHADTIGSG RAIPIKQGM L LKRS GKWLKTWKKKYVTLC SNGMLTYY SSLGDYMKNIHKKEIDLQTSTIKVPGKWPSLATSACT PISSSKSNGLSKDMDTGLGDSICFSPSISSTTSPKLN PPPSPHANKKKHLKKKSTNNFMIVSATGQ TWHFEATT YEERDAWVQAIQSQILASLQSCSSKSKS QLT SQSEA MALQSIQNMRGNAHCVD CETQNP KWASNLGVLMCIE CSGIHRSLGPHLSRVRSL ELDDWPVELRKVMSSIVND LANSIWEGSSQGQTKPSEKSTREEKERWIRSKYE EKL FLAPLPCTELSLGQQLLRATADEDLQTAILLLAHGSC EEVNETCGEGDGCTALHLACRKG NVVLAQLLI WYGVD VMARDAHGNTALTYARQASSQECINVLLQY GCPDECV *YLFYLTAVSLVQKQNGKNKDNSEFQKEITNSANNSI FSTFRKLSKYTKC
2220	A	1	1594	NGGGSLNNYSSPIPSTPAPSR RDPQFSVPPTANTPTP VCKRSMRWSNLFTSEKGS DPKERKAPENHADTIGSG RAIPIKQGM L LKRS GKWLKTWKKKYVTLC SNGMLTYY SSLGDYMKNIHKKEIDLQTSTIKVPGKWPSLATSACT PISSSKSNGLSKDMDTGLGDSICFSPSISSTTSPKLN PPPSPHANKKKHLKKKSTNNFMIVSATGQ TWHFEATT YEERDAWVQAIQSQILASLQSCSSKSKS QLT SQSEA MALQSIQNMRGNAHCVD CETQNP KWASNLGVLMCIE CSGIHRSLGPHLSRVRSL ELDDWPVELRKVMSSIVND LANSIWEGSSQGQTKPSEKSTREEKERWIRSKYE EKL FLAPLPCTELSLGQQLLRATADEDLQTAILLLAHGSC EEVNETCGEGDGCTALHLACRKG NVVLAQLLI WYGVD VMARDAHGNTALTYARQASSQECINVLLQY GCPDECV *YLFYLTAVSLVQKQNGKNKDNSEFQKEITNSANNSI FSTFRKLSKYTKC
2221	A	1	1594	NGGGSLNNYSSPIPSTPAPSR RDPQFSVPPTANTPTP VCKRSMRWSNLFTSEKGS DPKERKAPENHADTIGSG RAIPIKQGM L LKRS GKWLKTWKKKYVTLC SNGMLTYY SSLGDYMKNIHKKEIDLQTSTIKVPGKWPSLATSACT PISSSKSNGLSKDMDTGLGDSICFSPSISSTTSPKLN PPPSPHANKKKHLKKKSTNNFMIVSATGQ TWHFEATT YEERDAWVQAIQSQILASLQSCSSKSKS QLT SQSEA MALQSIQNMRGNAHCVD CETQNP KWASNLGVLMCIE CSGIHRSLGPHLSRVRSL ELDDWPVELRKVMSSIVND LANSIWEGSSQGQTKPSEKSTREEKERWIRSKYE EKL FLAPLPCTELSLGQQLLRATADEDLQTAILLLAHGSC EEVNETCGEGDGCTALHLACRKG NVVLAQLLI WYGVD VMARDAHGNTALTYARQASSQECINVLLQY GCPDECV *YLFYLTAVSLVQKQNGKNKDNSEFQKEITNSANNSI FSTFRKLSKYTKC
2222	A	1	1594	NGGGSLNNYSSPIPSTPAPSR RDPQFSVPPTANTPTP VCKRSMRWSNLFTSEKGS DPKERKAPENHADTIGSG RAIPIKQGM L LKRS GKWLKTWKKKYVTLC SNGMLTYY

643
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SSLGDYMKNIHKKEIDLQTSTIKVPGKWPSLATSACT PISSSKSNGLSKDMDTGLGDSICFSPSISSTTSPKLN PPSPHANKKKHLKKKSTNNFMIVSATGQTHFEATT YEERDAWVQAIQSQILASLQSCSSKSKSQTLSQSEA MALQSIQNMGRNAHCVDCEQNPKNWASLNLGVLMCIE CSGIHRSLSLPHLSRVRSLELDDWPVELRKVMSSIVND LANSIWEGSSQGQTKPSEKSTREEKERWIRSKYEELK FLAPLPCTELSLGQQLLRATADEDLQTAILLLAHGSC EEVNETCGEGDGCTALHLACRKGNVVLAQLLIWYGV VMARDAHGNTALTARQASSQECINVLLQYGCPCDEC *YLFYLTAVSLVQKQNGKNKDNSEFQKEITNSANNSI FSTFRKLSKYTKC
2223	A	2	415	GGFAAAVESFHHEDVLLFAALMAHELGHNLGIQHDHS ACFKDKHFCMHENITKESGFSSCSDYFYQFLREH KGACLFNKPRPRGRKRRDSACGNGVVEDTDQCDGSL CQHHACCDENCILKAKA*CNDGPCCHK
2224	A	53	325	MRLSVCLLLLTALCCYRANAVVCQALGSEITGFLLA GKPVFKFQLAKFKAPLEAVAAMKMEVKKCVDTMAYEKR VLITKTLGKIAEKCDR*
2225	A	9	422	ESRERSGNRRGAEDRGTCGLQSPSAMLGAKPHWLPGP LHSPGLPLVLVLLALGAGWAQEGSEPVLLGECELVVC EPGRAAAGGPGGAALGEAPPGRVAFAAVRSHHHEPAG ETGNGTSGAIYFDQVLVNEGGGFDRAS
2226	A	42	722	MGCDGRVSGLLRRNLQPTLTYSVFFSFGLCIAFLGP TLLDLRCQTHSSLPQISWVFFSQQLCLLLGSALGGVF KRTLAQSLWALFTSSLAISLVFAVIPFCRDVKVLASV MALAGLAMGCIDTVANMQLVRMYQKDSAVFLQVLHFF VGFGALLSPLIADPFLSEANCLPANSTGQHHLPRATC SMSPGCWGQHHVDAQALVQPDVPAKADSQGPGREPEGP MPSG*
2227	A	42	722	MGCDGRVSGLLRRNLQPTLTYSVFFSFGLCIAFLGP TLLDLRCQTHSSLPQISWVFFSQQLCLLLGSALGGVF KRTLAQSLWALFTSSLAISLVFAVIPFCRDVKVLASV MALAGLAMGCIDTVANMQLVRMYQKDSAVFLQVLHFF VGFGALLSPLIADPFLSEANCLPANSTGQHHLPRATC SMSPGCWGQHHVDAQALVQPDVPAKADSQGPGREPEGP MPSG*
2228	A	2	474	TGPTIKNMDGTFNVTSCCLKLNSSQEDPGTVYQCVVRH ASLHTPLRSNFTLTAAHSLSETEKTDNFSIHWWPIS FIGVGLVLLIVLIPWKKICNKSSSAYTPLKCILKHWN SFDQTQTLKKEHLIFFCTRAWPSYQLQDGEAWPPEGSV NINTYSTTV
2229	A	2	1654	GRGDSSSSSGSGSGSGSRACPARPSAPGLRAPTPPP RLPGASGAPAAARLTLLKFLAVLLAAGMLAFLGAVICII ASVPLAASPARALPGGADNASVASGAAASPGPQRSLS ALHGAGGSAGPPALPGAPAASAHPLPPGPLFSRFLCT PLAAACPSGAQQGDAAGAAPGEREELLLQSTAEQLR QTALQQEARIRADQDTIRELTGKLGRCESGLPRGLQG AGPRRDTMADGPWDSPALILELEDAVRALRDRIDRLE ELPARVNLSAAPAPVSAVPTGLHSMKMDQLEGQLLAQV

644
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LAL EKERVALSHSSRRQRQEVEKELDV LQGRVAELEH GSSAYSPPDAFKISIPIRNNMYARVRKALPELYAFT ACMWLRSRSGTGQGT PFSYSVPG/QAGNEIVLLEAG HEPMELLINDKVAQLPLSLKDNGWHHICIAWTTRDGL WSAYQDGELQSGENLAAWHPKPHGILILGQEQDTL GGRFDATQAFVGDIAQFNLWDHALTPAQVLGIANCTA PLLGNVLPWEDKLVEAFGGATKA AFDVCKGRAKA
2230	A	3	913	FMTDVNSWLLTFGFQLHNVI PGYPKPDMDAMEPSYEL IHTQMKTQEWDNSKSILGVQCEVQQLKAFVTLERFD QLYGSTITSCQQAPKTKKFASSGSVFGKGVKFALKDG RVTTDIISVANEDGRRVAAILNHAHYLENLHFTIDGV DTHYFVKPGPSEGDLAILGLSGGRRTLENGVNVTVSQ INTVLNGRTRRYTDIQLQYGALCLNTRYGTTLDEEKA RVLELSRQRAVRQAWAREQQRLREGEEGLRAWTEGEK QQVLSTGRVQGYDGFVVISVEQYPELSDSANNIHFM RQSEMGR
2231	A	488	75	ASVPKTNKIEPRYSIIIPSCGIQAARACFEHSNFFKV NASGPAGHSAKSIEGAPRGKGRGRAVARLAADRPPAP KIQLRAF*LQQL*YTLLELELPRLAPDLPSNGSSLK DLKWTHSNYRASKECIVIFRHYLP
2232	A	3	181	HERDVLFNLCENLVKSSEANS PAHEEFKTMLLIAHYY ATRSAAESVYQL*AVSRVLLSLVY
2233	A	1	492	KIKAKNLTNYDLCSIFLGTSTLLVWVG VIRYLG YFQA YNVLILTMQASLPKVLRF CACAGMIYLG YTF CGWIVL GPYHDKFENLNTVAECLFSLVNGDDMFATFAQIQQKS ILVWLF SRLYLYSFI SLFIYMILSLFIALITDSYDTI KKFQQNGFPETDLQEF
2234	A	1	492	KIKAKNLTNYDLCSIFLGTSTLLVWVG VIRYLG YFQA YNVLILTMQASLPKVLRF CACAGMIYLG YTF CGWIVL GPYHDKFENLNTVAECLFSLVNGDDMFATFAQIQQKS ILVWLF SRLYLYSFI SLFIYMILSLFIALITDSYDTI KKFQQNGFPETDLQEF
2235	A	1	576	PCGEFHHSS/QKATPAEEVEDSNDSSYSEPPDVQQQL NHYSQAALARNNSRVSPVPLSGAAAGTEQKTEAVLHC EFCEFSSGYIQSIRRH YRDKHGGKKLFKCKDCSFYTG FKSAFTMHVEAGHSAPVEEGPKDLRCPLCLYHTKYKR NMIDHIVLHREERVPIEVCRSKLSKYLQGVVFRCDK CTFTCSR
2236	C	60	472	MPLLEYARNMLRTWSSLPWTRFRVCLLSLSLFLWANR LED SRSCQPNPMSLTTLPGHRLKEAVWLPAPSR TMSP HLDPNQLGILLRVLRKEKEDGDYPDMATHPSSRYEA CSSGITLAAPP THGPRPTDPRIGPAP
2237	C	60	472	MPLLEYARNMLRTWSSLPWTRFRVCLLSLSLFLWANR LED SRSCQPNPMSLTTLPGHRLKEAVWLPAPSR TMSP HLDPNQLGILLRVLRKEKEDGDYPDMATHPSSRYEA CSSGITLAAPP THGPRPTDPRIGPAP
2238	A	129	329	VSNIVDPHQTVGLSTQEPGDI FTYSEFDGILGLAYPS LASE*SVPVLDNTMQRHLVAQDLFSVYMSR
2239	A	130	502	DSRI PKEAPDQQKKMGPPSLVLCLLSATVFSLLGGS SAFLSHHRLKGRFQDRRNIRPNII LVLTDDQDVELG

645
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SMQVMNKTRRIMEQGGAHFINAFVTTMCCPSRSSIL TGKYVHNHNTYMY
2240	A	3	498	YKEVVTQHFL*VTYETHPIYYLKISQPSGNPKKIIWM DCGIHAREWIAPAFQWFVKEILQNHKDNSRIRKLLM NLDFYVLPVLNIDGYIYTWTDLWRKSRSPHNNGTC FGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPE TKAVASFIESKNDDFCA
2241	A	3	498	YKEVVTQHFL*VTYETHPIYYLKISQPSGNPKKIIWM DCGIHAREWIAPAFQWFVKEILQNHKDNSRIRKLLM NLDFYVLPVLNIDGYIYTWTDLWRKSRSPHNNGTC FGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPE TKAVASFIESKNDDFCA
2242	A	972	468	MAAAGAGRLRRVASALLLRSPRLPARELSAPARLYHK KVVDHYENPRNVGSLDKTSKNVGTGLVGAPACGDVMK LQIQVDEKGKIVDARFKTFGCGSAIASSSLATEWVKG KTVEEALTIKNTDIAKELCLPPVKLHCSMLAEDAIKA ALADYKLKQEPKKGEAEKK
2243	A	1193	548	TQAWTRAEKDRKGSVRALRLHLERGPPPT*RGSHPL/Q SVPCIQKPSIFSSYP/ GLPQSGGEPGPVGEQQPVRR PEQPSCGPASRMPLTSRSVPPGRGALPPDSLSTRKGL PRPSTAGHRVRESGHKVPVSQRLNLPVMGATRSLNLP PRKVAVPGPTR*RDQDSKQDFSSKPLQSVPGLASTQQ TLTPADSGPGTGGRDATRAGLPGVETMGNGVD
2244	A	3	773	SLEMSSDGEPLSRMDSSEDSISSTIMDV DSTISSGRST PAMMNGQGSTTSSSKNIAYNCCWDQCQACFNSSPDLA DHRSIHVDGQRGGVFVCLWKGCKVYNTPTSTSQSWLQ RHMLTHSGDKPFKCVVGGCNASFASQGGLARHVP THF SQQNSSKVSSQPKAKEESPSKAGMNKRKLNKRRRS LARPHDFFDAQTLDAIRHRAICFNLSAHIESLGKHS VVFHSTVSILLFFQIKYKTLQKNISTIIISKSLKI
2245	A	3834	2068	GARGRPLAETWPF LTAPVLPGLQITEPTMAEKGDCI ASVYGYDLGGRFVDFQPLGFGVNGLVLSA VDSRACRK VAVKKIALSDARSMKHALREIKIIRRLDHDNIVKVYE VLGPKGTDLQGE LFKFSVAYIVQEYMETDLARLLEQG TLAEHAKLFMYQLLRGLKYIHSANVLHRDLKPANIF ISTEDLV LKIGDFGLARIVDQHYS\ HKGYLSEGLVTK WYRSPRLLLS PNNYTKAIDMWAAGCILAEMLTGRMLF AGAHELEQMQLILETIPVIREEDKDELLRVMPSFVSS TWEVKRPLRKLLPEVNSE AIDFLEKILTFNPMDRLTA EMGLQHPYMSPYSCPEDEPT SQHPFRIEDEIDDIVLM AANQS QLSNWDTCSSRYPVSLSSDLEWRPDRCDASE VQRDPRAGSAPLAENVQVDPRKDSHSSSERFLEQSHS SMERAFEADYGRSCDYKVGSPSYLDKLLWRDNKPHHY SEPKLILDL SHWKQAAGAPPTATG\ LADTGAREDEPA SLFLE\ IAQWVKSTQG\ AQSTPARPPTTPSAACLPRP P\ PPGPGGCR\ RQPPVRPGRVHLPRPEALHQARGPAG Q
2246	A	328	595	VIEWVVPVEPPNQLSTSSVGRVPGSTRPQRSFLSRVV RAALPLQLLLLLLLLLLACL LPSSEEDYSCTQANNFAR SFYPMLRYTNGPPPT

646
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2247	A	548	811	SSFIKRHILIFEDDWHQTTTCHHPHPHP\F*RCQFHIF YVSVQNSISPSLSVSSSHPDPRPDHEVHQHRAAHHHQH GQGPLGHGLVARVG
2248	A	37	441	GXAGVGGDSEGEVTSALSATFSGPKIAFYVGLKSPHE GYEVLKFDDVVTNLGNHYDPTTGKFSCQVRGIYFFTY HILMRGGDGTSMWADLCKNGQVRASAIQDADQNYDY ASNSVVLHLDGDEVYVKLDGGKA
2249	A	808	112	RRYKSGTEVNNTDGGIARLIVFGTGQKDWATDPKEP ADLVAIAFGGVCVGFSSNAKFGHPNNIIGVGGAKSMAD GWETARRLDRPPILENDENGILLVPGCEWAVFRLAHP GVITRIEIDTKYFEGNAPDSCKVDGCILTQTQEEAVIR QKWILPAHKWKPLLPVTKLSPNQSHLFDSLTLLELQDV ITHARLTIVPDGGVNRLRLRGFPSSICLLRPREKPML KFSVSFKANP
2250	A	189	1811	PPFGGLSAAQTIGEMWEAQFLGLLFLQPLWVAPVKPL QPGAEPVWVAQEGAPALPCSPTIPLQDLSLLRRAG VTWQHQPDSGPPAAAPGHPLAPGPHPAAPSSWGPRPR RYTVLSVGPGLRSGRLPLQPRVQLDERGRQRGDFSL WLRPARRADAGEYRAAVHLRDRALSCRLRLRLGQASM TASPPGSLRASDWVILNCSFSRPDRPASVHWFRNRGQ GRVPVRESPPHHLAESFLFLPQVSPMDSGPWGCILTY RDGFNVSIMYNLTVLGLEPPTPLTVYAGAGSRVGLPC RLPAGVGTRSFLLTAKWTPPGGGPDLLVTGDNGDFTLR LEDVSQAQAGTYTCHIHLQEQQLNATVTLAIITVTPK SFGSPGSLGKLLCEVTPVSGQERFVWSSLDTPSQRSF SGPWLEAQEAQLLSQPWQCQLYQGERLLGAAVYFTL SSPGAQRSGRAPGALPAGHLLFLTLGVLSLLLLVTG TFGFHLWRRQCRP\RRFSALEQGIH\P\RQAQSKIEE LEQEPEPEPEPEPEPEPEPEPEPEQL
2251	A	3	3773	SWPRGRGETGGHPGALRTRTMQKSVRYNEGHALYLAF LARKEGTRKGFSLSKKTAEASRWHEKWFALYQNVLFYF EGEQSCRPAAGMYLLEGCSCECTPAPPRAGAGQGGVRD ALDKQYYFTVLFHGHEGQKPLELRCEEEQDGKEWMEAI HQASYADILIEREVLMOQYIHLVQIVETEKIAANQLR HQLEDQDTEIERLKESEIINALNKTKERMRPYQSNQED DPDIKKIKKVQSFMRGWLCCRKWKTIQDYICSPHAE SMRKRNRQIVFTMVEAESEYVHQLYILVNGFLRPLRMA ASSKKPPISHDDVSSIIFLNSETIMFLHEIFHQGLKAR IANWPTLILADLFDILLPMLNIYQEFVRNHQYSLQVL ANCKQNRDFDKLLKQYEANPACEGRMLETFLTYPMFQ IPRYIITLHELLAHTPHEHVERKSLEFAKSKLEELSR VMHDEVSDTENIRKNLAIERMIVEGCDILLDTSQTFI RQGSLLIQVPSVERGKLSKVRLGSLSLKKEGERQCFLE TKHFLICTRSSGGKLHLLKTGGVLSLIDCTLIEEPDA SDDDSKSGSQVFGHLDFKIVVEPPDRAAFTVLLAPS RQEKAAWMSDISQCVDNIRCNGMLTIVFEENSKVTVP HMIKSDARLHKDDTDICFSKTLNSCKVPQIRYASVER LLERLTDLRFLSIDFLNTFLHTYRIFTTAAVVLGKLS DIYKRPFTSIPVRSLELFFATSQNNRGEHLVDGKSPR LCRKFSPPPLAVSRTSSPVRARKLSLTSPLNSKIGA

647
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LDLTTSSSPTTTTQSPAASPPPHGTGQIPLDLSRGLSS PEQSPGTVEENVNDNPRVDLCNKLKRSIQKAVLESAPA DRAGVESSPAADTTTELSPCRSPSTPRHLRYRQPGGQT ADNAHCSVSPASAFAIATAAAGHGSPPGFNNTERTCD KEFTIIRRTATNRVLNVLRLHWVSKHAQDFELNNELKMN VLNLLEEVLRLDPDLLPQERKAAANILMALSQDDQDDI HLKLEDIIQMTDCMKAECFESLSAMELAEQITLLDHV IFRSIPYEEFLGQGWMKLDKNERTPYIMKTSQHFNNDM SNLVASQIMNYADVSSRANAIEKWVAVADICRCLHNY NGVLEITSALNRSAYRLKKTWAKVSKQTKALMDKLQ KTVSSEGRFKNLRETLKNCNPPAVPYLGMYLTDLAFI EEGTPNFTEEGLVNFSSKMRMISHIIREIRQFQQTSYR IDHQPKVAQYLLDKDLIIDEDTLYELSLKIEPRLPA
2252	A	1	4602	ASGNLKNARFSAIYRQDSNKLSDNDMLKLLADFRKP EKMAKLPVILGNLDITIDNVSSDFPNYVNSSYIPTKQ FETCSKTPITFEVEEFVPCIPKHTQPYTIYTNHLYVY PKYLKYDSQKSFARKNIAICIEFKDSDEEDSQPLKC IYGRPGGPVFTRSAFAAVLHHHQNPFEYDEIKIELPT QLHEKHHLLLTFFHVSCDNSSKSGSTKKRDVETQVGY SWLPLLKDGRRVVTSEQHIPVSANLPSGYLGYQELGMG RHYGPEIKWVDGGKPLLKISTHLVSTVYTQDQHLHNF FQYCQKTESGAQALGNELVKYLKSLHAMEGHVMI AFL PTILNQLFRVLTRATQEEVAVNVTRVIIHVVAQCHEE GLESHLRSYVKYAYKAEPYVASEYKTVHEELTKSMTT ILKPSADFLTSNKLLKYSWFFFDVLIKSMASHLIENS KVKLLRNQRFPAHYHHAETVNVNMLMPHITQKFRDNP EASKNANHSLAVFIKRCFTFMDRGFVFKQINNYISCF APGDPKTLFEYKFEFLRVVCNHEHYIPLNLPMPFGKG RIQRYQDLQLDYSLTDEFRCRNHFLVGLLLREVGTALQ EFREVRLIAISVLKNLLIKHSFDDRYASRSHQARIAT LYLPLFGLLIENVQRINVRDVSPFPVNAGMTVKDESL ALPAVNPLVTPQKGSTLDNSLHKDLLGAISGIASPYT TSTPNINSVRNADSRGSLISTDSGNSLPERNSEKSN LDKHQQSSTLGNSSVVRCDKLDQSEIKSLLMCFLYILK SMSDDALFTYWNKASTSELMDFFTISEVCLHQFQYMG KRYIARTGMMHARLQQLGSLDNLSTFNHSGHSDADV LHQSLLLEANIATEVCLTALDTLSLFTLAFKNQLLADH GHNPLMKKVFDVYLCFLQKHQSETALKNVFTALRS LI YKFPSTFYEGRADMCALCYEILKCCNSKLSSIRTEA SLLLYFLMRNNFDYTGKKSFVRTHLQVII SVS QLIAD VVGIGGTRFQQSLSIINNCANSRDLIKHTSFSSDVKD LTKRIRTVLMATAQMKHEHNDPEMLVDLQYSLAKSYA STPELRKTWLD SMARIHVKNLSEAMCYVHV TALV AEYLTRKEAVQWEPPLPHSHSACLRRSRGGVFRQGC TAFRVITPNIDEEASMMEDVGMQDVHFNEDVLMELLE QCADGLWKAERYELIADIYKLIIPYIEKRRDFERLAH LYDTLHRAYSKVTEVMHSGRLLGTYFRVAFFGQAAQ YQFTDSETDVEGFFEDGKEYIYKEPKLTPLSEISQ RLKLYSDKFGSENVKMIQDSGKVNPKDLDSKYAYIQ VTHVIPFFDEKELQERKTEFERSHNIRRFMFEMPFTQ

648
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TGKRQGGVEEQCKRRTILTAIHCFPYVKKRIPVMYQH HTDLNPIEVAIDEMSKKVAELRQLCSSAEVDMIKLQL KLQGSVSVQVNAGPLAYARAFLLDDTNTKRYPDNKVKL LKEVFRQFVEACGQALAVNERLIKEDQLEYQEEMKAN YREMAKELSEIMHEQLG
2253	A	1	782	MRMEAGEAAPPAGAGGRAAGGWGKWVRLNVGGTVFLT TRQTLCREQKSFLSRLCQGEELQSDRDETGayLIDRD PTYFGPIILNFLRHGKLVLDKDMAEEGVLEEAFFYNIG PLIRIIKDRMEEKDYTVTQVPPKHVYRVLCQEEELT QMVSTMSDGWRFEQLVNIGSSYNYGSEDQAEFLCVVS KELHSTPNGLSSESSRKTSTEEQLEEQQQQEEVEE VEVEQVQVEADAQEK/CCYKPEAPGCEAPDHLQGLGV PI
2254	A	2407	2216	SGCVEMLYSHSLEYNPEWISVQSAVAPAQLALNSDGD L*LHSGERTRRD*QLPEAGGPGLQEPLQLGELDITSD EFILDEV DG\VDLRHYSKQVELELQQIEQKSIRDYIQ ESENIALSHNQITACDAVLERMEQMLGAFQSDLSSIS SEIRTLQEQSGAMNIRLRNRQAVRGKLGELVDGLVVP SALVTAILEAPVTEPRFLEQLQELDAKAAAVREGEAR GTAACADVRGVLDRLRVKAVTKIREFILQKIYSFRKP MTNYQIPQTALLKYRFFYQFLLGNERATAKEIRDEYV ETLSKIYLSYYRSYLGRMLKVQYEEVAEKDDLGMVED TAKKGFFSKPSLRSRNTIFTLGTRGSVISPTLEAPI LVPHTAQRGEQRYPFALFRSQHYALLDNSCREYLF CEFFVVS GPAHDLFHVMGRTLSMTLKHLD SYLADC YDAIAVFLCIHIVLRFRNIAAKRDVPALDRYWEQVLA LLWPRFELILEMNVQSVRSTDPQRLGGLDTRPHYITR RYAEFSSALVSINQTI PNERTMQLLGQLQVEVENFVL RVAAEFSSRKEQLVFLINNYDMMLGVLM\E*ERAADD SKEVESFQQLLNARTQEFIEELLSPPFGLVAFVKEA EALIERGQAERLRGEEARVTQLIRGFGSSWKSSVESL SQDVMRSFTNFRNGTSIIQGALTQLIQ\LYHRFHRV\ LSQPQLRALPARAELINIHHLMVLELKKHKPNF
2255	A	1205	462	ASITVSSGRIPTSLSVGPPGAPLHRPQKPREGAWDME DVAPTGVRQAFSELFPFPHVLPEPGFPDTPSQVYSP GLPPAPAQPSSI PPCALVSQPTVQFILQGSPLVGC AAQTLAPVPAALTPASEPASQATAASNSEEKTPAPRL AAEKTKEEYMKKLHMQERAVEEVKLAIKPFYQKREV TKEEYKDILRKAVQKICHSKSGEINPVKVANLVKAYV DKYRHMRRHKKPEAGEEPPTQGAEG
2256	A	1205	462	ASITVSSGRIPTSLSVGPPGAPLHRPQKPREGAWDME DVAPTGVRQAFSELFPFPHVLPEPGFPDTPSQVYSP GLPPAPAQPSSI PPCALVSQPTVQFILQGSPLVGC AAQTLAPVPAALTPASEPASQATAASNSEEKTPAPRL AAEKTKEEYMKKLHMQERAVEEVKLAIKPFYQKREV TKEEYKDILRKAVQKICHSKSGEINPVKVANLVKAYV DKYRHMRRHKKPEAGEEPPTQGAEG
2257	A	901	521	FFFGNGVSPCRQAGV*WHDLDLQNLPPGFKRFSYLS LPSSW\DYRHVLP RQANFCIF/M*RRGFTMLARMVSI S*PRDLPALASQSAGITGVSHHAPPQMDFTFALLCFA

649
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LKGCLPRQKEGGTLNLI
2258	A	186	1338	TRMSRHEGVSCDACLKGNFRGRRYKCLICYDYDLCAS CYESGATTTRHTTDHPMQCILTRVDFDLYYGGEAFSV EQPQSFTCPYCGKMGYTETSLQEHVTSEHAETSTEVI CPICAALPGGDPNHVTDDFAAHLTLEHRAPRDLDESS GVRHVRRMFHPGRGLGGPRARRSNMHFTSSSTGGLSS SQSSYSPSNREAMDPIAELLSQLSGVRRSAGGQLNSS GPSASQLQQLQMLQLERQHAQAARQOLETARNATR TNTSSVTTTITQSTATTNIANTESSQOTLQNSQFLLT RLNDPKMSETERQSMESERADRSFLVQELLSTLVRE ESSSDEDDRGEMADFGAMGCVDIMPLDVALENLNLK ESNKGNEPPPPPL
2259	A	1157	481	SWPGQAEPSEREFVREAAETRGSEVFEIMNPVYSPG SSGVPIYANAKGIGYPAGFPMGYAAAAPAYSPNMPGA NPTFQTGYTPGTPYKVSCSPTSGAVPPYSSSPNPYQT AVYPVRSAYPQQSPYAQQGTYYTQPLYAAPPHVIHHT TVVQPNGMPATVYPAPIPPPRGNGVTMGMVAGTTMAM SAGTLLTAHSPTPVAPHPVTVPTYRA\QGTPTYSYVP PQW
2260	A	33	563	MVLSVPVIALGATLGTATSILALCGVTCLCRHMHPKK GLLPRDQDPDLEKAKPSLLGSAQRFNKKSTEPVQPR ALLKFPDIYGPRPAVTAPENVINYADYSRSTEEPTAP ASPQPPNDSRLKRQVTEELFILPQNGVVEDVCVMTW NPQKAGSWNQAPKLHYCLDYDCHKAEC*
2261	A	6120	2968	HPSPGFDRVRAAMPNTIIEALRGTMDPALREAAERQ LNEAHKSLNFVSTLLQITMSEQLDLPVRQAGVIYLKN MITQYWPDRRETAPGDISPYTIPEEDRHCIRENIVEAI IHSPELIRVQLTTCIHIIKHDPSPRWTAIVDKIGFY LQSDNSACWLGILLCLYQLVKNYEYKKPEERSPLVAA MQHFLPVLKDRFIQLLSQSDQSVLIQKQIFKIFYAL VQYTLPLELINQONLTIEWIEILKTVVNRDVPNETLQV EEDDRPELPWWKCKKVALHILARLFERYGSPGNVSKE YNEFAEVFLKAFVGVQQVLLKVLYQYKEKQYMAPRV LQOTLNYINQGVSHALTWKNLKPFIQGIQDVI FPLM CYTDADEELWQEDPYEYIRMKFDVFEDFISPTTAAQT LLFTACSKRKEVLQKTMGFCYQILTEPNADPRKKDGA LHMIGSLAEILLKKKI\YKDQMEYMLPESMYSPLF\S SELG\YMRARACWVLHYFCEVKFKSDQNLQTALELTR RCLIDDRMPVKVEAAIALQVLISNQEKAKEYITPFI RPVMQALLHIIRETENDDLTNVIQKMICEYSEEVTPI AVEMTQHLAMTFNQVIQTGPDEEGSDDKAVTAMGILN TIDTLLSVVEDHKEITQQLEGICLOVIGTVLQQHVLE FYEEIFSLAHSITCQQVSPQMWQLLPLVFEVFQQDGF DYFTDMMPLLNHYVTVDLTLSDTKYLEMIYSMCKK VLTGVAGEDAECHAAKLLEVIILQCKGRGIDQCIPLF VEAALERLTREVKTSSEL*TMGLQVAIAALHYNAYLLL NTLENLHFPNNVEPVTNHFI/QWLNDVDCFLGLHDDR MCVLSLCALIDMEQIPQGLNQVSGQILPAFILLFNGL KRAYACHAEHENDSDDDDAEEDDDETEELGSDEDDID EDGQEYLEILAKQAGEDGDDDEDWEEDDAEETALEGYS

650
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				TIIDDEDNPVDEYQIFKAI FQTIQNRNPVWYQALTHG LNEEQRKQLQDIATLADQRRAAHESKMIEKHGGYKFS AP\VPSSSFNFGGPAPGMN
2262	A	13	2237	AEGCAERRGTEPVVELSMSWESGAGPGLGSQGM DLVW SAWYGKCVKGKGS LPLSAHGI VVAWLSRAEWDQVT VY LFCDDHKLQRYALNRITVWRSRSGNELPLAVASTADL IRCKLLDVTGGLGTDELRLLYGMALVRFVNLI SERKT KFAKVPLKCLAQEVNIPDWIVDLRHELTHKKMPHIND CRRGCYFVLDWLQKTYWCRQLENSLRETWELEEFREG IEEEDQEEDKNIVVDDITEQKPEPQDDGKSTESDVKA DGDSKGSEEVDSHCKKALSHKELYERARELLVS YEEE QFTVLEKFRYLPKAIKAWNNPSPRVEC VLAELKGVTC ENREAVLDAFLDDGFLVPTFEQLAALQIEYEENVDLN DVLVPKPFSSQFWQPLLRLGLHSQNFTQALLERMLSELP ALGISGIRPTYILRWTVELIVANTKTGRNARRFSAGQ WEARRGWRLFNCASASLDWPRMVESCLGSPCWAS PQLL RIIF\KAMGQGLQDE\EQEKLLRICSIYTQSGENSLV QEGSEASPIGKSPYTLDSLYWSVKPASSSFGSEAKAQ QQEEQGSVNDVKEEKEEKEVLPDQVEEEEENDDQEE EEDEDEDEDEDEEDRMEVGPFSTGQESPTAENARLLA QKRGALQGS AWQVSSSEDVRWDTFP\LGRMPRSRPRTP AELMLENYDTHVIFWTKPVL\EQRLEPSTCK\TDTLG L\SCGVGS\GNCSNSSSSNFRGAFLLEARGSLH\GL\ KTGLQLF
2263	A	1	528	LGNTVLHYCSMYSKPECLKLLLRSKPTVDIVNQAGET ALDIAKRLKATQCEDLLSQAKSGKFNPHVHVEYEWNL RQEEIDESDDDLDDKPSPVKKERSPRPQS FCHSSSIS PQDKLALPGFSTPRDKQRLSYGAFTNQIFVSTSTDSP TSPTTEAPPLPPRNAGKGPTGPPI TPHR
2264	A	422	2	APGASVGRAQAAG*RGGPTGRPPSALGVS/EAGRAG RAGEGRPVPPAYPLCKSAQTS GPPKARLS\ PPLASCG GRGPPGGAACATCAPPAGPARSSRCRRRS PPE*GPR* PSRPARPSPGSAASRRQKLTPCRCQFRGLCA
2265	A	1	1742	VSAVEFVLHGKDFQVDCKASGSPVP*ISWSLLDGTMI NNAMQADDSGHRTRRYTLFNNGTLYFNKVGVAEEGDY TCYAQNTLGKDEMKVHLTVITAAPRI RQSNKTNKRIK AGDTAVLDC EVTGD PKPKIFWLLPSNDMISFSIDRYT FHANGSLTINKVKLLDSGEYVCVARNPSGDDTKMYKL DVVSKPPLINGLYTNRTVIKATAVRHSHKHFDCRAEG TPSPEVMWIMPDNIFLTAPYYGSRITVHKNGTLEIRN VRLSDSADFICVARNEGGE SVLVVQLEVLEMLRRPTF RNPFFNEKIVAQLGKSTALNCSVDGNPPPEI IWILPNG TRFSNGPQSYQYLIASNGSFIISKTTREDAGKYRCAA RNKVGYYIEKLVILEIGQKPVILTYAPGT VKGISGESL SLHCVSDGIPKPNIKWTMPSGYVVD RPQINGKYILHD NGTLVIKEATAYDRGNYICKAQNSVGHTLITVPVMIV AYPPRITNRPPRSIVTRTGAA FQLHCVALGV PKPEIT WEMPDHSL LSTASKERTHGSEQLHLQGT LVIQNPQTS DSGIYKCTAKNPLGSDYAATYIQVI
2266	A	2334	68	RWHQAPGPVRQRPPDDLQPGPGL\WMPGPARTTESA

651
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GQKIKELLSGIGNISERVSLKK/RG*PAEQGTDKPQ RGHERE\RAL\RQAARAPDPSAPPAPPGAACRHRMC SWPPAC*RTAAASAWTRGGPCASRCSPAPVLSTWTRP APPSPTSPWTSSAPVLGGR*RYPWALVRTAGSTCP* PCPA\PVLQSQGGGGRGPCPLRL*G/PPFWMSAPPTS CPSKR\GLPAPEQAHSQHAASALPWGPATHHTGPLP TRPHPRPWGHFSGNLSGAWQPASRTRLPAGRVPAPIC GFHQGVGGA/GSELP*RTATQACPCAVPPCSGSLLRM LLWTS*GPEHYLPSR\DGP*WRQSPHRPRG/VP*PT CAQQGPSRPWRFKWKAP\SGRHLQGAPCRAHADDDG DRAGRPGQLRS*SPCAVPPDPQPRDTAAGGADPAR PALHGG*GQLLCHRPEAATGVPAAPPQHPAVTRRA CPWALATLPASVTAPPGLMG*RETELAWPEPSGKVG GHVGAERS*KCLEAVEHKADSDWEQPRRALNLAGRSF ASSAGVSPSLTAAAPAL/GLPHCWAAFPFPQQLRP GGSAGHSGPGGP\GNRISGVWTWGEFVTVAATPPGAP AAPLGGTTRCPTVPLSHCSH\CPAAHSGTPR\WRVLP ETKAQNSMQGAPASARGLVPHQGRASGWPVAGMLNN* VPPAGAVPSTVHYFQGHSG\GAVAGGGP*APAPSLLP QPG\HGPPPGAGVFIWGGCSRRSRRCRHCPR
2267	A	29	175	KSRPGTVAHACNPSTLGSRRGRIIPAQEFKTSLGNTV SE\PCLYLRKNN
2268	A	29	175	KSRPGTVAHACNPSTLGSRRGRIIPAQEFKTSLGNTV SE\PCLYLRKNN
2269	A	961	365	PRVRLNGCGRALALGRGLKSFLRGTSLEEIMSLALR SELVVDKTKRKKRRELSEEQKQEI KDAFELFDTDKDE AIDYHELKVAMRALGFDVKKADVLLKILKDYDREATGK ITFEDFNEVVTDWILERDPHEEILKAFKLFDDDDSGK ISLRNLRRVARELGENMSDEELRAMIEEFDKDGDGEI NQEEFIAIMTGDI
2270	A	131	1567	NKLVTERQILGDPTYMRQADGRKVLRSSIREFLCSEA MFHLGVPTRAGACVTSESTVVRDVFYDGLDPLRFLS LQMSTQGVQAPAW/RRNDIRVQLLDYVISSFYPEIQ AHASDSVQRNAAFFREVTRRTARMVAEWQCVGFCHGV LNTDNMSILGLTIDYGPFGFLDRYDPDHVCNASDNTG RYAYSKQPEVCRWNLRKLAEALQPELPLELGEAILAE EFDAEFQRHYLQKMRRKLGLVQVELEEDGALVSKLLE TMHLTGADFTNTFYLLSSFPVELESPGLAEFLARLME QCASLEELRLAFRPQMDPRQLSMMLMLAQSNPQLFAL MGTRAGIARELERVEQQSRLEQLSAAELQSRNQGHWA DWLQAYRARLDKDLGAGDAAWQAEHVRVMHANNPK YVLRNYIAQNAIEAAERGDFSEVRRVLKLETPYHCE AGAATDAEATEADGADGRQRSYSSKPPLWAAELCVT* SSFYPEIQAAHASDSVQRNAAFFREVTRRTARMVAEW QCVGFCHGV LNTDNMSILGLTIDYGPFGFLDRYDPDH VCNASDNTGRYAYSKQPEVCRWNLRKLAEALQPELPL ELGEAILAEFEFQAEFQRHYLQKMRRKLGLVQVELEED GALVSKLLETMHLTGADFTNTFYLLSSFPVELESPGL AEFLARLMEQCASLEELRLAFRPQMDPRQLSMMLMLA QSNPQLFALMGTRAGIARELERVEQQSRLEQLSAAEL

652
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				QSRNQGHWADWLQAYRARLDKDLEGAGDAAAWQAEHV RVMHANNPKYVLRNYIAQNAIEAAERGDFSEVRRVLK LLETPTYHCEAGAATDAEATEADGADGRQRSYSSKPPL WAAELCVT
2271	A	131	1567	NKLVTERQILGDPTYMRQADGRKVLRSSIREFLCSEA MFHLGVPTRAGACVTSESTVVRDVFYDGLDPLRFLS LQMSTQGVQAPAW/RRNDIRVQLLDYVISSFYPEIQA AHASDSVQRNAAFFREVTRRTARMVAEWQCVGFCHGV LNTDNMSILGLTIDYGPFGLDRYDPDHVCNASDNTG RYAYSKQPEVCRWNLRLKLAELQPELPLELGEATLAE EFDAEFQRHYLQKMRRKLGLVQVELEEDGALVSKLLE TMHLTGADFTNTFYLLSSFPVELESPGLAEFLARLME QCASLEELRLAFRPQMDPRQLSMMLMLAQSNPQLFAL MGTRAGIARELERVEQQSRLEQLSAAELQSRNQGHW ADWLQAYRARLDKDLEGAGDAAAWQAEHVRVMHANNPK YVLRNYIAQNAIEAAERGDFSEVRRVLKLETPYHCE AGAATDAEATEADGADGRQRSYSSKPPLWAAELCVT* SSFYPEIQAHAASDSVQRNAAFFREVTRRTARMVAEW QCVGFCHGVNLNTDNMSILGLTIDYGPFGLDRYDPDH VCNASDNTGRYAYSKQPEVCRWNLRLKLAELQPELPL ELGEAILAEFEFQAEFQRHYLQKMRRKLGLVQVELEED GALVSKLLETMTHLTGADFTNTFYLLSSFPVELESPGL AEFLARLMEQCASLEELRLAFRPQMDPRQLSMMLMLA QSNPQLFALMGTRAGIARELERVEQQSRLEQLSAAEL QSRNQGHWADWLQAYRARLDKDLEGAGDAAAWQAEHV RVMHANNPKYVLRNYIAQNAIEAAERGDFSEVRRVLK LLETPTYHCEAGAATDAEATEADGADGRQRSYSSKPPL WAAELCVT
2272	A	53	439	FFLPLLIICYIFIFRAMRETGRALQTFGACKGNGE SLWQRQLQSECKMAKIMLLVILLFVLSWAPYSAVAL VAFAGYAHVLTPYMSSVPAVIAKASAIHNPIIYAITH PKYRVAIAQHLPLCLGVLL
2273	A	9	410	MTTTFPPRKMVAQFLLVAGNVANITTVSLWEEFSSSD LADLRFLLDMSQNQFQYLPDGFRLKMPSLSHLNLHQC LMTLHIREHEPPGALTELDLSHNQLSELHLAPGLASC LGSLRLFNLSNQLLGVPVPGPLY
2274	A	73	489	FLLLRASPEHTCVKSKTLDPMVIFFTSGTTGFPKMA KSHHGLALQPSFPGSRKLRLSLKTSVSWCLSDSGWIV ATIWTLVPEPWTAGCTVFIHHLPPQFDTKVIIQTLVKYP INHFWGVSSIYRMILQQDFTSIRFPAL
2275	A	3	1238	LTKMHLTENPHPQVTHVSSSQSGCSIASDSGSSSLSD IYQATESEVGVDLTRLPEGPVDSDEDEEEDDEIDRT DPLQGRDLVRECLEKEPADKTDDDIEQLLEFMHQLPA FANMTMSVRRELCSVMIFEVVEQAGAIILEDGQELDS WYVILNGTVEISHPDGKVENLFMGNSFGITPTLDKQY MHGIVRTKVDDCQFVCIAQQDYWRILNHVEKNTHKVE EEGEIVMVHEHRELDRLSGTRKGHIVIKATPERLIMHL IEEHSIVDPTYIEDFLTYRTFLESPLDVGIKLLEWF KIDSLRDKVTRIVLLWVNNHFNDFEGDPAMTRFLEEF EKNLEDTKMNGHLRLNLNIAACAkakWRQVVLQKASRE

653
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SPLQFSLNGGSEKGFIFVEGVEPGSKAADSGLKRGD QIMEV
2276	A	3	1238	LTKMHLTENPHQVTHVSSSQSGCSIASDSGSSSLSD IYQATESEVGVDLTRLPEGPVDSEDEEEDEEIDRT DPLQGRDLVRECLEKEPADKTDDIEQLLEFMHQLPA FANMTMSVRRELCSVMIFEVVEQAGAIILEDGQELDS WYVILNGTVEISHPDGKVENLFMGNSFGITPTLDKQY MHGIVRTKVDDCQFVCIAQQDYWRILNHVEKNTHKVE EEEGIVMVHEHRELDMSGTRKGHIVIKATPERLIMHL IEEHSIVDPTYIEDFLLTYRTFLESPLDVGIKLEWF KIDSLRDKVTRIVLLWVNNHFNDFEGDPAMTRFLEEF EKNLEDTKMNGHLRLNLNIACAAKAKWRQVVLQKASRE SPLQFSLNGGSEKGFIFVEGVEPGSKAADSGLKRGD QIMEV
2277	A	1	794	FRGFDRGDCAALPCTYPHSPCSH*GGNCLPSLLTRP CVKA*PQMSGRKSSMRWRQRSLTAGTSS*TPTSST MC*ALVGSSTWNCMLQAGSTAPGAGTPGSRPTWSSSS TCSWTAPSGRARCACASSSSCAMSAARRGWTSPACWR RTSRAWTTSSPACASSATASVAASTASTWPAARTTG GTAESSARPARRASCTGSPARSCWRRRRPPTPSPGRP APPSRRTTRAQAGTSALSPGACFGPRSCC*SSTCSSL SVAPY
2278	A	269	832	MGSSRLAALLPLLLIVIDLSDSAGIGFRHLPHWNTR CPLASHTDDSFSGSSAYIPCRTWWALFSTKPCVVRVW HCSRCLCQHLLSGGSGLQRLFLHLLVQKSKKSSTFKF YRRHKMPAPAQRKLLPRRHLSEKSHHISIPSPDISHK GLRSKRTPPFGSRDMGKAFPKWDSPTPGGDRPSSFEL LP*
2279	A	269	832	MGSSRLAALLPLLLIVIDLSDSAGIGFRHLPHWNTR CPLASHTDDSFSGSSAYIPCRTWWALFSTKPCVVRVW HCSRCLCQHLLSGGSGLQRLFLHLLVQKSKKSSTFKF YRRHKMPAPAQRKLLPRRHLSEKSHHISIPSPDISHK GLRSKRTPPFGSRDMGKAFPKWDSPTPGGDRPSSFEL LP*
2280	A	2	381	VLPTAQGKLYQDDLKVN PANVSHLVSPFTWQGPGGHL KAPQWTTSSLFPFQIRNVGTGLCADTKHGALGSPLRL EGCVRG\RGEAAWNNMQVRAAPQGLAARFSETSAAWG ADTASWEGEAWVSDK
2281	A	1	993	MRDLFGTRLRRAEDVFPPVIGVAAHKGGVYKTSVSVH LAQDLALKGLRVLLVEGNDPQGTASMYHGWVPDLHIH AEDTLLPFYLGEKDDVTYAIKPTCWPGLDIIPSCAL HRIETELMGKFDEGKLPTDPHMLRLAIETVAHDYDV IVIDSAPNLGIGTINVVCAADVLIPTPAELFDYTSA LQFFDMLRDLLKNVDLKGFEPPDDLKKSFKSPEPRFT PEEFFRIIFNRSIDAFKDFVVASETSDCVVSSTLSPEK VLRASWKRSDNSLSKLSPTQIRLGEVLTPVMSAFWE AEVWNSGDSDDMALDFDCTSSEVDAESTNRKVLRP
2282	A	3	582	SLYQFSVVETAGPGLTVGRLRAQDPDLGDNALMAYSI LDGEGSEAFSISTDLQGRDGLLTVRKPLDFESQRSYS FRVEATNTLIDPAYLRRGPFKDVASVRVAVQDAPEPP

654
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				AFTQAAYHLLTVPENKAPGTLVGQISAADLDSPASPIR YSILPHSDPERCFISIQPEEGTIHTAAPLDREARAWN LTVLATEL
2283	A	3	582	SLYQFSVVEVETAGPGLVGRLLRAQDPDLGDNALMAYSI LDGEGSEAFSISTDLQGRDGLLTVRKPLDFESQRSYS FRVEATNTLIDPAYLRRGPFKDVASVRVAVQDAPEPP AFTQAAYHLLTVPENKAPGTLVGQISAADLDSPASPIR YSILPHSDPERCFISIQPEEGTIHTAAPLDREARAWN LTVLATEL
2284	A	1	831	KNVWKRWKKRFFVLVQVIQYTFAMCSYREKKAEPQEL LQLDGYTVDYTDPPQGLEGGRAFFNAVKEGDTVIFAS DDEQDRILWVQAMYRATGQSHKVPPTQVQKLNAGG NVPQLDAPISQFYADRAQKHGMDEFISSNPCNFDHAS LFEMVQRLTLDHRLNDSYSLGWFSFGQVFLDEYCA RNGVRGCHRHLCYLRDLLERAENGAMIDPTLXHYSFA FCASHVHGHRPDGIGNC*LLKKRNVF*RKSKEEXSXV LLRKIRLQHFRXLLFPFG
2285	A	140	445	MQPSGLEPGTFGRWPLLSLLLLLLLLLQPVTCAYTTP GPPRALTTLGAPRAHTMPGTYPSTTLSSPSTQGLQE QARALMRDFPLVDGHNDLPLVLRQVYHN
2286	A	294	1568	MSLTIWTVCGVLSLFGALSYAELGTTIKKSGGHYTYI LEVFGPLPAFVRVWVELLIIRPAATAVISLAFGRYIL EPFFIQCEIPELAIKLITAVGITVVMVLNSMSVSWSA RIQIFLTFCKLTAILIIIVPGVMQLIKGQTQNFKDAF SGRDSSITRLPLAFYYGMYAYAGWFYLNFTVEVENP EKTIPLAICISMAIVTIGYVLTNVAYFTTINAEELL SNAVAVTFSERLLGNFSLAVPIFVALSCFGSMNGGVF AVSRLFYVASREGLHPEILSMIHVRKHTPLPAVIVLH PLTMIMLFSGDLDSLNLNFLSFARWLFGLAVAGLIYL RYKCPDMHRPFKVPFLFIPALFSFTCLFMVALSLYSDP FSTGIGFVITLTGVPAYYLFIIWDKKPRWFRIMSEKI TRTLQIILEVVPEDKL*
2287	A	3397	630	SPGGRTPAARDSVREVIQNSKEVSIYVWQEKNCAS SAVRCKLSRRGDGQA*C*EINQ\NLAEAEAGLNITH\I CLA\PDSSEAEIIDEILKINEDTRVHGLALQISENLF SNKVLNALKPEKDVGDVTDINLGKLVGRDAHECFVSP VAKAVIELLEKSGVNLDGKKILVVGAGHSLEAALQCL FQRKGSMTMSIQWKTRQLQSKLHEADIVVLGSPKPEE IPLTWIQPGTTVLNCSHDFLSGKVGCGSPRIHFGLI EEDDVILLAAALRIQNMVSSGRRWLREQQHRRWRLHC LKLQPLSPVPSDIEISRGQTPKAVDVLAKEIGLLADE IEIYGKSKAKVRLSVLERLKDQADGKYVLVAGITPTP LGEGKSTVTIGLVQALTAHLNVNSFACLRQPSQGPTF GVKGGAAAGGYAQVIMPEEFNLHLTGDIHAITAANNL LAAAIIDTRILHENTQTDKALYNRLVPLVNGVREFSEI QLARLKKLGINKTDPSTLTETEEVSKFARLDIDPSTIT WQRVLDTNDRLRKITIGQGNTEKGHYRQAQFDIAVA SEIMAVLALTDLSADMKARLGRMVVASDKSGQPVTAD DLGVTGALTVLMKDAIKPNLMQTLEGTPVFVHAGPFA NIAHGNSSVLADKIALKLVGEEGFVVTEAGFGADIGM

655
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EKFFNIKCRASGLVPNVVVLVATVRALKMHGGGPSVT AGVPLKKEYTEENIQLVADGCCNLQKQIQITQLFGVP VVVALNVFKTDTRAIEDLVCELAKRAGAFDAVPCYHW SVGGKGSVDLARAVREAASKRSRFQFLYDVQVPIVDK IRTIAQAVYGAKDIELSPEAQAKIDRYTQQGFGNLPI CMAKTHLSLSHQPKKGVPRDFILPISDVASIGAGF IYPLVGTMTMPGLPTRPCFYDIDLDTETEQQVKGLF
2288	A	474	4247	IIISIIISTSNKIKMSEAPRFFVGPEDTEINPGNYRHFF HHADEDEEEDDSPPERQIVVGICSMAKKSQIPNPMK EILERISLFKYITLVVFEEVILNEPVENWPLCDCLI SFHSKGFPLDKAVAYAKLRNPFVINDLNMQYLIQDRR EVYSILQAEGILLPRYAILNRDPNNPKECNLIEGEDH VEVNGEVFQKPFVEKPVSAEDHNVYIYYPTSAGGGSQ RLFRKIGSRSSVSPESNVRKTGSYIYEEFMPTDGTD VKVYTVGPDYAHAEARKSPALDGKVERDSEGKEVRYP VILNAREKLIWVKVCLAFKQTVCGFDLLRANGQSYVC DVNGFSFVKNSMKYYDDCAKILGNIVMRELAPQFHIP WSIPLAEDIPIVPTTSGTMMELRCVIAVIRHGDRTF KQKMKMEVRHQKFFDLFEKCDGYKSGKLKPKKPKQLQ EVLDIARQLLMELGQNNNDSEIEENKPKLEQLKTVLEM YGHFSGINRKVQLTYLPHGCPKTSSEEDSRREEPSL LLVLKWGGELTPAGRVQAEELGRAFCMPYGGQGDYA GFPGCGLLRHLSTYRHDLDKIYASDEGRVQMTAAFAK GLLALEGELTPILVQMVKSANMNGLLSDSDSLSSCQ QRVKARLHEILQKDRDFTAEDYEKLTPSGSISLIKSM HLIKNPVKTCDKVYSLIQSLTSQIRHRMEDPKSSDIQ LYHSETLELMLRWSKLEKDFKTKNGRYDISKIPDIY DCIKYDVQHNGFLEIRKTQWELYRLSKALADIVIPQE YGITKAEKLEIAKGYCTPLVRKIRSDLQRTQDDDTVN KLHPVYSRGVLSPERHVRTRLYFTSESHVHSLLSILR YGALCNESKDEQWKRAMDYLVNVNELNYMTQIVIMLY EDPNKDLSSSEERFHVLELHFSPGAKGCEEDKNLPSGYG YRPASRENEGRRPFKIDNDDEPHTSKRDEVDRVILF KPMVSEPIHIHRKSPLPRSRKTATNDEESPLSVSSPE GTGTWLHYTSGVGTGRRRRRSGEQITSSPVSPKSLAF TSSIFGSWQQVVSENANYLRTPTLVEQKQNPTVGSH CAGLFSTSVLGGSSSAPNLQDYARTHRRKLTSSGCID DATRGSAVKRFYISFARHPTNGFELYSMVPSICPLET LHNALSLKQVDEFLASIASPSSDVPRKTAEISSALR SSPIMRKKVSLNTYTPAKILPTPPATLKSTKASSKPA TSGPSSAVVPNTSSRKKNITSKTETHEHKKNTGKKK
2289	A	3	552	FIDDELA TEWSLTMETLT KVLARNLYSLDLSDLPLDK LSEQQKKHKKGKGVGHEFQKVSVDKSF SRGWSRDQPG QAPMRQRSATTTGSPGTEKARSIVRQKTVDIDDAQIL PRSTRVRHFSQSEETGNEVFGALNEEQPLPRSSSTD ILEPFTVERAKGAVPVIDSSSRHAPSLQSFTEASS
2290	A	3	147	QPLNHYFICSSHNTYLVGDQLCGQSSVEGYIRCSGGR EGVQLMRGTM
2291	B	1	498	MDLCQKNETDLENAENNEIQFTEETEPTYTCPDGKSE KNHVYCLLDVSDITLEQDEKAKEFIIGTGWEEAPPQR

656
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SSPAVGLRQPGLPGPHLLGPTGGRKGLGGTRHQGP EQRNAFGTAWTPETHPTRGHTGRTEAAVAGGDARPEG RLLGSRQLNRLPDAETQ
2292	A	963	5	LDFLCHRDMGDNITSITEFLLLGFVGPRIQMLLFGL FSLFYVFTLLGNGTILGLISLDSRLHAPMYFFLSHL\ AVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTF FSTFAVTECLLLVVMYSYDLYV\AICHPLRYLAIMTWR VCITLAVTSWTTGVLLSLIHLVLLPLPFCRPQKIYH FFCEILAVLKLACADTHINENMVLGAI SGLVGPLST IVVSYMCILCAILQIQSREVQRKAFCTCFSHLCVIGL FYGTAIMYVGPRYGNPKEQKKYLLLFHSLFNPM LNP LICSLRNSEVKNTLKRVLGVERAL
2293	A	1306	158	ISYCPKF PNRDQRDKDGDGVGDACDSCPDVSNPNQSD VDNDLVGDS CDTNQSDGDGHQDSTDNCPTVINSAQL DTDKDGIGDECDDDDNDGIPDLVPPGPDNCRLVPNP AQEDSNSDGVGDICESDFDQDQVIDRIDVCPENAEVT LTDFRAYQTVVLDPEGDAQIDPNWVVLNQGMEIVQTM NSDPGLAVGYTAF\NGVDFEGTFHVNTQTDDDYAGFI FGYQDSSSFYVVMWKQTEQTYWQATPFRAVAEPGIQL KAVKSKTGPGEHLRNSLWHTGDTSDQVRLLWKDSRNV GWKDKVSYRWFLQHRPQVGYIRVRFYEGSELVADSGV TIDTTMRGGRLGVFCFSQENI IWSNLKYRCNDTIPED FQEFQTQNFDRFDN
2294	A	4701	866	DAPGRPPVRLPTMELEDGVVYQEEPGGSGAVMSERVS GLAGSIYREFERLIVRYDEEVVKELIPLVVAVLENLD SVFAQDQEHQVELELLRDDNEQLITQYEREKALRKHA EEKFIEFEDSQEQEKKDLQTRVESLESQTRQLELKAK NYADQISILEEREAE LKKEYNALHQRHTEMIHNYMEH LERTKLHQLSGSDQLESTAHSRIRKERPISLGIFPLP AGDGLLTPDAQKGGETPGSEQWKFOELSQPRSHTSK DELSDV SQGSKATTPASTANS DVATIPDTPLKEEN EGFVKVTDAPNKSEISKHIEVQVAQETRNVTGSAEN EEKSEVQAI IESTPELMDMDKDL SGYKGSSTPTKGIEN KAFDRNTESLFEELSSAGSGLIGDVDEGADLLGMGRE VENLILENTQLLETKNALNIVKNDLIAKVDELTCCKD VLQGELEAVKQAKLKLEEK NRELEEEELRKARAE AEDA RQKAKDDDDSDIPTAQRKRFRTRVEMARVLMERNQYKE RLMELQEAVRWTE MIRASREN PAMQEKKRSSIWQFFS RLFSSSSNTTKKPEPPVNLKYNAPTSHVTPSVKKRSS TLSQLPGDKSKAFDFLSEETEASLASRREQKREQYRQ VKAHVQKEDGRVQAFGWSLPQKYKQVTNGQGENKMN LPVPVYLRPLDKKDTSMKLWCAVG VNL SGGKTRD GGS VVGASVFYKDVAGLDTEGSKQRSASQSSLDKLDQELK EQQKELKNQEELSSLVWICTSTHSATKVLIIIDAVQPG NILDSTVCNSHVLCIASVPGARETDYPAGEDLSESG QVDKASLCGSMTSNSSAETDSLLGGITVVGCSAEGVT GAATSPSTNGASPVMDKPPEMEAEENSEVDENVPTAEE \ATEATEGNAGSAEDTV\DISQTGVYTEHVFTDPLG\ VQIPEDLSPVYQSSNDS DAYKDQISVLPNEQDLVREE AQKMSSLLPTMWLGAQNGCLYVHSSVAQWRKCLHSIK

657
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LKDSILSIVHVKGIVLVALADGTLAIFHRGVDGQWDL SNYHLLDLGRPHHSIRCMTVVHDKVWCGYRNKIYVVQ PKAMKIEKSFDAHPRKESQVRQLAWVGDGVVWSIRLD STLRLYHAHTYQHLQDVDIEPYVSKMLGTGKLGFSFV RITALMVSCNRLWVG TGNGV I I S I P L T E T V I L H Q G R L LGLRANKTSGVPGNRPGSVIRVYGDENSDKVTPTGTFI PYCSMAHAQLCFHGH R D A V K F F V A V P G Q V I S P Q S S S S GTDLTGDKGRGHLHRS LVVRRP
2295	A	1	1668	AAAAAAGAFAGRRAACGAVLLTELLERAAFYGITSNL VLFLNGAPFCWEGAQASEALLLFMGLTYLGSPFGGWL ADARLGRARAILLSLALYLLGMLAFPLLAAPATRAAL CGSARLLNCTAPGPDAAARCCSPATFAGLVLVGLGVA TVKANITPFGADQVKDRGPEATRFFNWFYWSINLGA ILSLGGIAYIQQNVSFVTGYAIP TVCVGLAFVVF LCG QSVFITKPPDGSAFTDMFKILTYSCCSQKRSGERQSN GEGIGVFQQSSKQSLFDSCKMSHGGPFTEEKVEDVKA LVKIVPVFLALIPYWTVYFQMOTTYVLQSLHLRIPEI SNITTTPHTLPAAWLTMFDAVLILLIPLKDKLVDP LRRHGLLPSSLKRIAVGMFFVMCSAFAAGILESKRLN LVKEKTINQ T I G N V V Y H A A D L S L W W Q V P Q Y L L I G I S E I F A S I A G L E F A Y S A A P K S M Q S A I M G L F F F F S G V G S F V G S G L L A L V S I K A I G W M S S H T D F G N I N G C Y L N Y F F L L A A I Q G A T L L L F L I I S V K Y D H H R D H Q R S R A N G V P T S R R A
2296	A	132	695	TQRAATPLPNSPQEAAILGSRRNQAGRVREKVYRSLP GPAFLGESWKRLSVLQESFSLTTPRQSQMRKSDIFPK SLPSQFFGSFGKPVACVTCACSLQLLKFIPEKSDIDL LVYRIDHYQQRQLQALFFKKKFQERLAEAKPKVEGRAE GCRRLRVESYLIMILEKHFPDILNMPSELQHLPEAAK VK
2297	A	5	505	CKKCQKKFSSGYQLILHHRVHVIERPYECKECGKNFR SGYQLTLHQRFHTGEKPYECTECGKNFRSGYQLTVHQ RFHTGEKTYECTQCGKAFIYASHIAQHERIHTGGKPY ECQECGRAFSQGGHLRIHQRVHTGEKPYKCKECKGTF STRSXLVEHGRVHTDEKPY
2298	A	102	449	PAPASGFTQTWGDACDPAAPQRPLEACFSVQSRTSSP MEPPI PQSAPLTPNSVMVQPLLD SRMSHSRLQHPLTI LPIDQVK T S H V E N D Y I D N P S L A L T T G P K R T R G G A P E L A P T P A
2299	A	402	2624	MAESRGRLYLWMCLAALASFLMGFMVGWFIKPLKET TTSVRYHQSI RWKLVSEMKAENIKSFLRSFTKLPHLA GTEQNFL LAKKIQTQWKKGFLDSAKLVHYDVLLSYPN ETNANYISIVDEHETE I F K T S Y L E P P P D G Y E N V T N I V PPYNAFSAQGMPEGDLVYVNYARTEDFFKLEREMGIN CTGKIV I A R Y G K I F R G N K V K N A M L A G A I G I I L Y S D P A DYFAPEVQPYPKGWNLPGTAAQRGNVLNLNGAGDPLT PGYPAKEYTFR LDVEEGVGI PRIPVHPIGYND AEILL RYLGGIAPPDKSWKGALNVSYSIGPGFTGSDSFRKVR MHVYNINKITRIYNVVG T I R G S V E P D R Y V I L G G H R D S WVFGAIDPTSGVAVLQEIARSFGKLMSKGWRPRRTII

658
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				FASWDAEEFGLLGSTEWAEENVKILQERSIAYINSDS SIEGNYTLRVDCPTLLYQLVYKLTKEIPSPDDGFESK FLYESWVEKDPSPENKNLPRINKLGGSGDFEAYFQRL GIASGRARYTKNKKTDKYSSYPVYHTIYETFELVEKF YDPTFKKQLSVAQLRGALVYELVDSKIIIPFNIQDYAE ALKNYAASIYNLSKKHDQQLTDHGVSFDSLFSVAVKNF SEAASDFHKRLIQVDLNNPIAVRMMNDQLMLLERAFI DPLGLPGKLFYRHIIIFAPSSHKNKYAGESFPGIYDAIF DIENKANSRLAWKEVKKHISIAAFTIQAAAGTLKEVL *
2300	A	74	520	PGVGPCLSVPPSAPSLVFRSVAGGAGMAERGLEPSPA AVAALPPEVRAQLAELELELSEGDITQKGYEKKRSL LSPYSPQTQETDSAVQKELRNQTPAPSAAQTSAPSKY HRTRSGGARDERYRSGEELQNGQLNRFNSSMNCVS
2301	A	6256	5813	MALQLWALTLLGLLGAGASLRPRKLDFFRSEKELNHL AVDEASGVVYLGAVALYQLDAKLQLEQQVATGPVLD NKKCTPPIEASQCHEAEMTDNVNQLLLVDPKRRLVE CGQLLKGILRSARPEQHLPPPVLGRGRQGEVFRGQQ*
2302	A	402	578	MPTYWLANLRPGLOPFLHFLLEWLAVFCCKIMVLAA AGLLPTLHMASFFSNALYNCFY
2303	A	186	1338	TRMSRHEGVSCDACLKGNFRGRRYKCLICYDYDLCAS CYESGATTTRHTTDPHMQCILTTRVDFDLYYGGEAFSV EQPQSFTCPYCGKMGYTETSLQEHVTSEHAETSTEVI CPICAALPGGDPNHVTDDFAAHLTLEHRAPRDLDESS GVRHVRRMFHPGRGLGGPRARRSNMHFTSSSTGGLSS SQSSYSPSNREAMDPIAELLSQLSGVRRSAGGQLNSS GPSASQLQQLQMLQLERQHAQAARQQLETARNATTR TNTSSVTTTITQSTATTNIANTESSQOTLQNSQFLLT RLNDPKMSETERQSMESERADRSFVQELLLSTLVRE ESSSSDEDDRGEMADFGAMGCVDIMPLDVALENLNLK ESNKGNEPPPPPL
2304	A	126	397	PLTEDGSPGPPPEGFKDLRNQRPPPHGTGPWRGPGPSG PPRSGQVPDNSTRCFLSDFWSPQGDQRPSCPYTGARF RQGAAQHLRCPSRRRR
2305	A	3	457	RAFDVRRKSLRPCCPRDFHAGCLTVSGPSTVMGAVG ESLSVQCRYEEKYKTFNKYWCQPCLPWHMVETGG SEGVVRSDQVIITDHPGDLTFTVTLENLTADDAGKYR CGIATILQEDGLSGFLPDPFFQVQVLVSSASSTENS KTP
2306	A	1	1117	NSRVDDFVAVMAPRTLVLVLLSGALALTQTWAGSHSMR YFFTSVSRPGRGEPRFIAVG YVDDTQFVRFDSDAASQ RMEPRAPWIEQEGPEYWDGETRKVKASQTHRVDLGT LRGYYNQSEAGSHTVQRMVGCVDVGSWRFRLRGYHQYA YDGKDYIALKEDLRSWTAADMAAQTTKHKWEAAHVAE QLRAYLEGTCVEWLRRLRYLENGKETLQRTDAPKTHMTH HPI SDHEATLRCWALSFP AEITLTWQRDGEDQTQDT ELVETRPAGDGTQKWA AVVVP SGQEQR YTVCHVQHEG LPKPLTLRWEPSQPTIPIVGIIAGLVLF GAVITGAV VAAVMWRRKSSDRKGGSYSQAASSDSAQGS DVSLTAC KV

659
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2307	A	3	491	DAWVAHASGELPPQTTKTLARFIPEVAVAYPKSKPLT TQIKIKKPPKVTMKTGKSLLHLHSTLEMFAARWRSKA PMSLFLLEVHFNLKVQYSVHENQLQMATSDDRGN/Y TGFITSYLEEAYIPVVNDVLQVGLPLPDFLAMNYNLA ELDIVENALMLDLKLG
2308	A	3	491	DAWVAHASGELPPQTTKTLARFIPEVAVAYPKSKPLT TQIKIKKPPKVTMKTGKSLLHLHSTLEMFAARWRSKA PMSLFLLEVHFNLKVQYSVHENQLQMATSDDRGN/Y TGFITSYLEEAYIPVVNDVLQVGLPLPDFLAMNYNLA ELDIVENALMLDLKLG
2309	A	3	491	DAWVAHASGELPPQTTKTLARFIPEVAVAYPKSKPLT TQIKIKKPPKVTMKTGKSLLHLHSTLEMFAARWRSKA PMSLFLLEVHFNLKVQYSVHENQLQMATSDDRGN/Y TGFITSYLEEAYIPVVNDVLQVGLPLPDFLAMNYNLA ELDIVENALMLDLKLG
2310	A	3	491	DAWVAHASGELPPQTTKTLARFIPEVAVAYPKSKPLT TQIKIKKPPKVTMKTGKSLLHLHSTLEMFAARWRSKA PMSLFLLEVHFNLKVQYSVHENQLQMATSDDRGN/Y TGFITSYLEEAYIPVVNDVLQVGLPLPDFLAMNYNLA ELDIVENALMLDLKLG
2311	A	75	739	APRAAPRLTMVSRMVSTMLSGLLFWLASGWTPAFAYS PRTPDVSEADIQRLHGVMEQLGIARPRVEYPAHQ MNLVGPQSIEGGAHEGLQHLGPFGNIPNIVAEITGDN IPKDFSEDQGYDPNPNPCPVGKTADDGCLNTPTAE FSREFQLHQHLDPEHDYPGLGKWNKKLLYGKMGGE RRKRRSVNPYLQGGRLDNVVAKKSVPHFSDKDE
2312	A	2	606	PSIRKHGTHPFPPT*SSPSGSC\SHCIAHSQCRQSP HASC*RGSRWG*SGRAGWPAPGCR*AAPGLAGSAHPR PPPSNPRCPPDAGPPGSGDPGLAAPEPSNHGRQHTA AAAAAGESQRHGRPGLAA*QPPLDTGPAARGSPAPP GARPRGGGRQHRPQGLPQAQPQ*APGVRAAPRAAPP \GHAGPDQAPEKAARTRG
2313	A	42	706	PRGQMASTGLELLGMTLAVLGWLGLVSCALPLWKVT AFIGNSIVVAQVWVEGLWMSVQSTGQMCKVYDSL LALPQDLQAARALCVIALLLALLGLLVAITGAQCTTC VEDEGAKARIVLTAGVILLLAGILVLI PVCWTAHAI I QDFYNPLVAEALKRELGASLYLGWAAAALLMLGGGLL CCTCPPPQVERPRGPRLGYSIPSRSGASGLDKRDYV
2314	A	2	484	FVANMLCGLSRETPGEADDGPYSKGGKDAGADVCLA CRRQSIPEEFRGITVVELIKKEGSTLGLTISGGTDKD GKPRVSNLRPGGLAARSDLLNIGDYIRSVNGIHLTRL RHDEIITLLKNVGERVVL/EAPENNPRIISKTVDVSL YKEGNSFGFVLRGQ
2315	A	326	2002	GLSRMSTETELQVAVKTSAKKDSRKKGQDRSEATLIK RFKGEVRYKAKLIGIDEVSAARGDKLCQDSMMKLKG VVAGARSKGEHKQKIFLTISFGGIKIFDEKTGALQHH HAVHEISYIAKDITDHRAGYVCGKEGNHRFVAIKTA QAAEPVILDLRDLFQLIYELKQREELEKKAQKDKQCE QAVYQTILEEDVEDPVYQYIVFEAGHEPIRDPETEEN IYQVPTSQKKEGVYDVPKSPVSAVTQLELFGDMSTP

660
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				PDITSPPTPATPGDAFIPSSSQTLPASADVFSVPFG TAAVPSGYVAMGAVLPSFWGQQPLVQQQMVMGAQPPV AQVMPGAQPIAWGQPGLEFPATQQPWPTVAGQFPFAAF MPTQTMPLPAAMFQGPLTPLATVPGTSDSTRSSPQT DKPRQKMGKETFKDFQMAQPPVPSPRKPDPSTLTCT EAFSSYFNKVGVAQDTHDCDDFDISQLNLTPVTSTTP STNSPPTPAPRQSSPSKSSASHASDPTTDDIFEEGFE SPSKSEEQEAPDGSQASSNSDPFGEPSPGSGDNISP QGR
2316	A	132	428	VNVLNQEIEAFSLSEDTSSGLPEDRVSVSFRVLYPI VITSLGVFYDANDVGFORNITVKLYQAEQEEALFIAR FSPPSGCVQVNKLWYKPVEQFILPE
2317	A	2334	1226	TAAAPVAPGTMDATVLRKKGYIVGINLGKGSYAKVK SAYSERLKFNAVAVKIIARKKTPTDFVERFLPREMDIL ATVNHGSIKTYEIEFETSDGRIYIIMELGVQGDILLEF IKCQGALHEDVARKMFRQLSSAVKYCHDLDIVHRDLK CENLLLDKDFNIKLSDFGFSKRCLRDSNGRIILSKTF CGSAAYAAPVQLQSIPIYQPKVYDIWSLGVILYIMVCG SMPYDDSDIRKMLRIQKEHRVDFPRSKNLTCECKDLI YRMLQ\PDVS\KRLHIDEILSHSWLQPPKPK\ATSSA SFKREGEKGKYRAECKLDTKTGLRDPDHRPDHKLGAQTQ HRLLVVPENENRMEDRLAETSRAKDHHSAGAEVGKAS T
2318	A	993	848	TRYATPLAPGPGHPFSCSRMATHHTLWMGLALLGV GDLQAAPEAQVSVQPNFQQDKFLGRWFSAGLASNSSW LREKKAALSMCKSVVAPATDGGLNLTSTFLRKNQCET RTMLLQPAAGSLGSSYSRSPHWGSTYSVSVVETDYDQY ALLYSQGSKGPGEDFRMATLYSRTQTPRAELKEKFTA FCKAQGFTEDTIVFLPQTDKCMTEQ
2319	A	2	394	AIHVRCLLSPGH TAGHMSYFLWEDDCPDPPALFSGDA LSVAGCGSCLEGSAAQMYQSLAELGTLPPETKVFCGH EHTLSNLEFAQKVEPCNDHKRDEDDVPTVPSTLGEER LYNPFLRVAEEPVRKFTGKA
2320	A	2	762	LEEVLKSELSGNFECTALALLDHPSEYAAARQLQKAMK GLGTDESVLIEFLCTRINKEIIAIEKAYQRLFDRSLE SNVKGDTSGNLKKILVSLQANRNEGDDVDKDLAQD AKDLYDAGEGRWGTDLAFNEVLAKRSYKQLRATFQA YQILIGKDIEEAIEEETSGDLQKAYLTLVRCAQDCED YFAERLYKSMKGAGTDEETLIRIIVTRAEVDLQGIKA KFQEKYQKSLSDMVRSDTSGDFRKL LVAL LH
2321	A	3	1335	QHSSRAGISSVAMPWAPLGHSGSHQLCVTFSSLHCLT RRNMHQMTDGLDKPGQIRWPLAITLAIWILVYFCIW KGVGWTGKVYFSATYPYIMLIILFFRGVTLPGAKEG ILFYITPNFRKLSDSEVWLDAATQIFFSYGLGLGLSLI ALGSYNSFHNNVYRDSIIIVCCINSCTSMFAGFVIFSI VGFMHVTKRSIADVAASGPGLAFLAYPEAVTQLPIS PLWAILFFSMLLMGLGIDSQFCTVEGFITALVDEYPR LRNRRELFIAAVCIISYLI GLSNITQGGIYVFKLFDY YSASGMSLLFLVFFECVSI SWFYGVNRFYDNIQEMVG SRPCIWWKLCWSFFTPIIVAGVFIFSAVQMTPLTMGN

661
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				YVFPKWGQGVGWLMASSMVLI PGYMA YMFLT LKGS LKQRIQVMVQPS EDIVRPENGPEQPQAGSSTSKEAYI
2322	A	775	945	MMYILLVFLTLWLLIEMIHC LQNGDHRRT RPPTETGW LPLRFHLRTGKILRYLRGE*
2323	A	197	598	MSALRPLLLLLLPLCPGPGPGPGSEAKVTRSCAETRO VLGARGYSLNLI PPALISGEHLRVCPQEYTCCSSETE QRLIRETEATFRGLVEDSGSFLVHTLAARHRKFDEFF LEMLFETLAFFCPDLSSHSTGA*
2324	A	2031	56	GTAETFHSVHFCPPVPKAPESPSLDSALASPLDPQA LACTPASPPDSQPPASPDSEALDFETPSSSLAPQTP DSALASETLASPQSLPPASPLLEDREEGDLGKASELA ETPKEEKAEGAAMLELVGSILRGCVPGVYRVQTVPSA RRPVVKFCHRPSGLHGDVSLSNRLALHNSRFLSLCSE LDGRVRPLVYTLRCWAQGRGLSGSGPLLSNYALTLLV IYFLQTRDPPVLPTVSQLTQKAGEGEQVEVDGWDCSF PRDASRL EPSINVEPLSSLLAQFFSCVSCWDLRGSL LSLREGQALPVAGGLPSNLWEGRLGPLNLQDPFDLSH NVAANVTSRVAGRLQNC CRAAANYCRSLQYQRRSSRG RDWGLLPLLQPSSPSSLLSATPIPLPLAPFTQLTAAL VQVFREALGCHIEQATKRTRSEGGGTGESSQGGTSKR LKVDGQKNCCEEKGEEQQGCAGDGGEDRVEEMVIEVG EMVQDWAMQSPGQPGDLPLTTGKHGAPGEEGQPSHAA LAERGPKGHEAAQEWSQGEAGKGASLPSSASWRCALW HRVWQGRRRARRRLQQQTKEGAGGGAGTRAGWLATEA QVTQELKGLSGGEERPETEPLLSFVASVSPADRMLTV TPLQDPQGLFPDLHHFLQVFLPQAIRHLK
2325	A	3	262	SLSMCREVHVYEYIPSVRQTELCHYHELYYDAACTLG AYHPLL YEKLLVQRLNMGTQGD LHRKGKVVLPGFQAV HCPAPSPVI PHS
2326	A	241	1449	ASLCKGCFFVTHVLVII LPSLQSPPTFGFLLDIDGVL VRGHRVIPAALKAFRRLVNSQGQLRVPVVFVTNAGNI LQHSKAQELSALLGCEVDADQVILSHSPMKL FSEYHE KRMLVSGQGPVMENAQGLGFRNVVTVD ELRMAFPLLD MVDLERRLKTTP LPRNDFPRI EGVL L LGEPVRWETSL QLIMDVLLSNGSPGAGLATPPYPHLPVLASNMDLLWM AEAKMPRFGHGTFLLCLETIYQKVTGKELRYEGLMGK PSILTYQYAEDLIRRQAERRGWAAPIRKLYAVGDNPM SDVYGANLFHQYLQKATHDGAPELGAGGTRQQQPSAS QSCISILVCTGVYNPRNPQSTEPVLGGGEPFFHGHRD LCFSPGLMEASHVVNDVNEAVQLVFRKEGWALE
2327	A	241	1449	ASLCKGCFFVTHVLVII LPSLQSPPTFGFLLDIDGVL VRGHRVIPAALKAFRRLVNSQGQLRVPVVFVTNAGNI LQHSKAQELSALLGCEVDADQVILSHSPMKL FSEYHE KRMLVSGQGPVMENAQGLGFRNVVTVD ELRMAFPLLD MVDLERRLKTTP LPRNDFPRI EGVL L LGEPVRWETSL QLIMDVLLSNGSPGAGLATPPYPHLPVLASNMDLLWM AEAKMPRFGHGTFLLCLETIYQKVTGKELRYEGLMGK PSILTYQYAEDLIRRQAERRGWAAPIRKLYAVGDNPM SDVYGANLFHQYLQKATHDGAPELGAGGTRQQQPSAS QSCISILVCTGVYNPRNPQSTEPVLGGGEPFFHGHRD

662
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LCFSPGLMEASHVVNDVNEAVQLVFRKEGWALE
2328	A	1	359	ISGESIYWSQKPTPSSNASPWSEPAAVDVELTAYALL AQLTKPSLTQKEIAKATSIVAWLAKQRNAYGGFSSTQ DTVVALQALAKYATTAYVPSEEINLVVKSTENFQRTF NIQAVNRM
2329	A	1	359	ISGESIYWSQKPTPSSNASPWSEPAAVDVELTAYALL AQLTKPSLTQKEIAKATSIVAWLAKQRNAYGGFSSTQ DTVVALQALAKYATTAYVPSEEINLVVKSTENFQRTF NIQAVNRM
2330	A	1	359	ISGESIYWSQKPTPSSNASPWSEPAAVDVELTAYALL AQLTKPSLTQKEIAKATSIVAWLAKQRNAYGGFSSTQ DTVVALQALAKYATTAYVPSEEINLVVKSTENFQRTF NIQAVNRM
2331	A	1	359	ISGESIYWSQKPTPSSNASPWSEPAAVDVELTAYALL AQLTKPSLTQKEIAKATSIVAWLAKQRNAYGGFSSTQ DTVVALQALAKYATTAYVPSEEINLVVKSTENFQRTF NIQAVNRM
2332	A	1	359	ISGESIYWSQKPTPSSNASPWSEPAAVDVELTAYALL AQLTKPSLTQKEIAKATSIVAWLAKQRNAYGGFSSTQ DTVVALQALAKYATTAYVPSEEINLVVKSTENFQRTF NIQAVNRM
2333	A	21	446	MESAVRVESGVLVGVVCLLLACPATATGPEVAQPEVD TTLGRVRGRQVGKGTDRLVNVFLGIPFAQPPLGPDR FSAPHPAQPWEGVRDASTAPPMCLQDVESMNSSRFVL NGKQQIFSVSEDCLVLNVYSPA EVPAGSGRP
2334	A	320	171	AASTTDGSYKCLCLPGYVPSDKPNYCTPLNTALNLEK CPFGLPHLSGSS
2335	A	351	49	PASPPRWGCWGCWGRWDCFASRSPWARS*SRRPPRST AAAPRSPARPRTCAGCTRRTWKTGRPARSRRSGRTPR AGR*K*SPGSGTRTSRPGGRRRPAGAR
2336	A	3	813	THASENAHGQASSFANFLVRTYLKGDAGFDSEIFKRS TFGPSVEFTSVLKPVFAREKEPFSLSCLFSEDVLD AE SIQWFRDGSLLRSSRRRKILYTDRQASLKVSC TYKED EGLYMVRVPSPFGPREQSTYVLVRDAEAENPGAPGSP LNVRCLDVNRDCLILT WAPPSDTRGNPITAYTIERCQ GESGEWIACHEAPGGTCRCPIQGLVEGQSYRFRVRAI SRVGSSVPSKASELVVMGDHDAARRKTEIPFDLG NKI TISTDAFEDTV
2337	A	834	628	DIREYK*NNPLVHMRTDET*MTMK**MVKEKKIVKED WRKVHLAS*QSFPSFFVIEHSKAIRGSWFPQL
2338	A	834	628	DIREYK*NNPLVHMRTDET*MTMK**MVKEKKIVKED WRKVHLAS*QSFPSFFVIEHSKAIRGSWFPQL
2339	A	3	449	PGAPRVRLETHPEPLPSDTMVSSCCGSVCSDQGCGLE TCCRPSCCQTTCCRTTCCRPSCCVSSCCRPQCCQSV C CQPTCCRPSCCPSCCQTTCCRTTCCRPSCCVSSCCRP QCCQSVCCQPTCCRPSCSISCCRPSCCVSRCCRSQR C
2340	A	3	449	PGAPRVRLETHPEPLPSDTMVSSCCGSVCSDQGCGLE TCCRPSCCQTTCCRTTCCRPSCCVSSCCRPQCCQSV C CQPTCCRPSCCPSCCQTTCCRTTCCRPSCCVSSCCRP

663
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QCCQSVCCQPTCCRPSCSISSCCRPSCCVSRCCRSQR C
2341	A	3	449	PGAPRVRLLETHPEPLPSDTMVSSCCGSVCSDQGCGL TCCRPSCCQTTCRTTCCRPSCCVSSCCRPQCCQSV CQPTCCRPSCCPSCCQTTCRTTCCRPSCCVSSCCRP QCCQSVCCQPTCCRPSCSISSCCRPSCCVSRCCRSQR C
2342	A	38	1435	ACLICFRIGRGNC SRKICEEFLNPQILLTLELVVTLA GKNKCRCTMLETLSRQWIVSHRMEMWLLILVAYMFQ RNVNSVHMPTKAVDPEAFMNI SEIIQHOGYPCEEYEV ATEDGYILSVNRIPRGLVQPKKTGSRPVLLQHGLVG GASNWISNLPNNSLGFIADAGFDVWMGNSRGNASR KHKTLSDQDEFWAFSYDEMAREFDLPAVINFILQKTG QEKIYYVGYSQGTMTGFIAFSTMPELAQKIKMYFALA PIATVKHAKSPGTFKLLLPDMMIKGLFGKKEFLYQTR FLRQLVIYLCGQVILDQICSNIMLLGGFNTNMMNMS RASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGS ETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSN PEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHR MYNEIIHLMHQEETQPFPRTA
2343	A	38	1435	ACLICFRIGRGNC SRKICEEFLNPQILLTLELVVTLA GKNKCRCTMLETLSRQWIVSHRMEMWLLILVAYMFQ RNVNSVHMPTKAVDPEAFMNI SEIIQHOGYPCEEYEV ATEDGYILSVNRIPRGLVQPKKTGSRPVLLQHGLVG GASNWISNLPNNSLGFIADAGFDVWMGNSRGNASR KHKTLSDQDEFWAFSYDEMAREFDLPAVINFILQKTG QEKIYYVGYSQGTMTGFIAFSTMPELAQKIKMYFALA PIATVKHAKSPGTFKLLLPDMMIKGLFGKKEFLYQTR FLRQLVIYLCGQVILDQICSNIMLLGGFNTNMMNMS RASVYAAHTLAGTSVQNILHWSQAVNSGELRAFDWGS ETKNLEKCNQPTPVRYRVRDMTVPTAMWTGGQDWLSN PEDVKMLLSEVTNLIYHKNIPEWAHVDFIWGLDAPHR MYNEIIHLMHQEETQPFPRTA
2344	A	91	1042	VTMYKDCIESTGDYFLLCDAEGPWGIILES LAILGIV VTILLLLAFLEFLMRKIQDCSQWNVLPTQLLFLLSVLG LFGLAFAFIIELNQQTAPVRYFLFGVLFALCFSCLLA HASNLVKLVRGCVSFSWTTILCIAIGCSLLQII IATE YVTLMTRGMMFVNMTPCQLNVDFVLLVYVFLMAL TFFVSKATFCGPCENWKQHGR LIFITVLF SIIIWVW ISMLLRGNPQFQRQPQWDDPVVCIALVTNAWVFLLLY IVPELCILYRSCRQECPLQGNACPVTAYQHSFQVENQ ELSRDKWKVLLNSDFLSHSGA
2345	A	2	669	AHTMVPEEEEPQDREKGLWWVQVKVWSMAVVSILL SV CFTVSSVPHNFMYSKTVKRLSKLREYQQYHSSLTCV MEGKDIEDWSCCPTPWT SFQSSCYFISTGMQSWTKSQ KNCSVMGADLVVINTREEQDFIIQNLKRNSSYFLGLS DPGGRRHQWVDQTPYNEN\ SREYRMRFWHSGE PNNL DERCAI INFRSSEEWGWN DIHCHVPQKSI CKMKKIYI
2346	A	2	669	AHTMVPEEEEPQDREKGLWWVQVKVWSMAVVSILL SV CFTVSSVPHNFMYSKTVKRLSKLREYQQYHSSLTCV

664
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				MEGKDIEDWSCCPTPWTSFQSSCYFISTGMQSWTKSQ KNCSVMGADLVVINTREEQDFI IQNLKRNSSYFLGLS DPGRRHWQWVDQTPYNEN\ SREYRMRFWHSGEPPNL DERCAI INFRSSEEWGWN DIHCHVPQKSICKMKKIYI
2347	A	1	2093	MLVLNSWAQVIHWPQPPKVLGLQPLEKTQYGF LGTDR VEEKTSVITIRVSVTHRHNSYMEAE NLTELSKFLLLG LSDDPELQPVLFGLFLSMYLVTVLGNLLI ILAVSSDS HLHTPMYFFLSNLSFVDICFISTTVPKMLVSIQARSK DISYMGCLTQVYFLMMFAGMDTFLLAVMAYDRFVAIC HPLHYTVIMNPCLCGLLVLASWFI IFWFSLVHILLMK RLTFSTGTEIPHFFCEPAQVLKVACSNTLLNNIVLYV ATALLGVFPVAGILFSYSQIVSSLMGMSSTKGKYKAF STCGSHLCVVS LFGYGTGLGVYLSSAVTHSSQSSSTAS VMYAMVTPMLNPFIYSLRNKDVKGALERLLSRADSCL LRCPSYTEPQNLTGVSEFLLLGLSEDPELQPVLAGLF LSMYLVTVLGNLLI ILAVSSDSHLHTPMYFFLSNLSL ADIGFTSTTVPKMIVDMQTHSRVISYEGCLTQMSFFV LFACMDDMLLSVMAYDRFVAI CHPLHYRIIMNPRLCG FLILLSFFISLLDSQLHNLIMLQLTCFKDVIDSNFFC DPSQLLHLRCSDTFINEMVIYFMGAIFGCLPISGILF SYYKIVSPILRVPTSDGKYKAFSTCGSHLAVVCLFYG TGLVGYLSSAVLPSPRKSMVASVMYTVVTPMLNPFIY SLRNKDIQSALCRLHGRI IKSHHLHPFCYMG
2348	A	773	317	QCTQKAAEGYTQFYYVDVLDGKLACVNKCTKGTKSQM NCNLGTCQLQRS GPRCLCPNTNTHWYWG ETCFENIAK SLVYGIVGAVMAVLLALI ILI ILFSLSQ\ RKRHRPE SEGEADFGLENATNNG\ PTLETVD SGTELHIQ\ RPE MVASTV
2349	A	55	414	MALTGYSWLLLSATFLNVGA EISITL E PAQPSEGDNV TLVVHGLSGELLAYS WYAGPTLSVSYLVASYIVSTGD ETPGPAHTXREAVRPDGS LDIQGILPRHSSTYILQTF NRQLQTEVG
2350	A	1	790	RGYNPNVNAGIINSFATAAFRFGHTLINPILYRLNAT LGEISEGHLPPHKA LFSPSRI IKEGGIDPVL RGLFGV AAKWRAPSYLLSPELTQRLFS AAYSAAVDSAATI IQR GRDHGIPPYVDFRVFCNLTSVKNFEDLQNEIKDSEIR QKLRKLYGSPGDIDLWPALMVEDLIPGTRVGPTLMC/ ML/STQFQRLRDGDRFWYENPGVFTPAQLTQLKQASL SRVLC DNGDSIQQVQADV F/RKRQEYPQDYLNCKRES PNVDPAC
2351	A	1	790	RGYNPNVNAGIINSFATAAFRFGHTLINPILYRLNAT LGEISEGHLPPHKA LFSPSRI IKEGGIDPVL RGLFGV AAKWRAPSYLLSPELTQRLFS AAYSAAVDSAATI IQR GRDHGIPPYVDFRVFCNLTSVKNFEDLQNEIKDSEIR QKLRKLYGSPGDIDLWPALMVEDLIPGTRVGPTLMC/ ML/STQFQRLRDGDRFWYENPGVFTPAQLTQLKQASL SRVLC DNGDSIQQVQADV F/RKRQEYPQDYLNCKRES PNVDPAC
2352	A	1	671	NFLPRLLLLTGPPQVGKTGSYLQFLRILFRMLIRLLE VDVYDEEEINTDHNESSEVSQSEGEPPWDIESFSKMP

665
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				FDVSVHDPKYSLSLVYTEKLAGVKQEVIKESKVEEP RKRETVSIMLTKYAAYNTFHHCEQCRQYMDFTSASQM SDSTLHAFTFSSSMLGEEVQLYFIIPKSKESHFVFSK QGHLESMLPLVSDKNLNAVKSPIFTPSSG*HEHGH V
2353	A	2	805	RELHKEVEVAKRNLAQQKIISEMESKLVQQLAENK LLKEQENMKELVNNLLRMTQIKIDEKEQKSKDFLKAQ QKYTNIVKEMKAKDLEIRIHKKKKCEIYRRLREFAKL YDTIRNERNKFVNLLHKAHQVNEIKERHKMSLNELE ILRNSAVSQERKLQNSMLKHANNVTIRESMQNDVRKI VSKLQEMKEKKEAQLNNDRLANTITMIEEEMVQLRK RYEKAVQHRNERRGLSPGMITKDRFLPVYGEITTRNI QLEKKLMGL
2354	A	159	1028	MGLCVPPFAVTTSFSLGLEWDLNVRLHGQHLVQQLVL RTVRGYLETPOPEKALALSFHGWSGTGKNFVARMLVE NLYRDGLMSDCVRMFIAFHFPHPKYVDLYKEQLMSQ IRETQQLCHQTLFI FDEAEKLHPGLLEVLPGLHERRA PEGHRAESPWTIFLFLSNLRGDIINEVVLKLLKAGWS REEITMEHLEPHLQAEIVETIDNGFGHSRLVKENLID YFIPFLPLEYRHVRLCARDALFLSQELLYKEETLDEIA QMMVYVPKEEQLFSSQGCKSISQRINYFLS*
2355	A	736	17	*RAMNFSICFLEIGSI*TGRYCKTVLCKLRAVL*SFR VLNITKAYLVLFSSLYKNLICSSVRSVPLKKFLKSLS SILRDRFFK*T*NPRGERERVLLGDFE*DRFRKCLSL IPLGGECSDDLRTSPSLTALPPNSIHCCSDPCITSI NLEPIKLL*HLRPPEASTHEANFTMASPLFRPS*CFK KITPSTHKPEKKTRTSSSFTR*GKPRRNK*GFSAFNG LVFLGLKLPCPVPLV*NP
2356	A	506	1317	GRTSSGKAGMWKPGAESWPLHTGAAQVMWFELKYAGL QCVEKYLIYPVVNLALTVDVDAHTVVSHDPKYCFYCRA LLMTVAGLKLRLSAFCCPPQQYLTLAFTVLLFHFDP RLSQGFLLDYFLMSLLCSKLWDLLYKLRVLTYYIAPW QITWGSFAFHAFAPFAVPHSAMLVQALLSGLFSTPL NPLLGSVAFIMSYARPLKFWERDYNTKRVDHSNTRLV TQLDRNPGADDNNLNSIFYEHLTRSLQHTLCGDLVLG RWGNYGPGDCF
2357	A	506	1317	GRTSSGKAGMWKPGAESWPLHTGAAQVMWFELKYAGL QCVEKYLIYPVVNLALTVDVDAHTVVSHDPKYCFYCRA LLMTVAGLKLRLSAFCCPPQQYLTLAFTVLLFHFDP RLSQGFLLDYFLMSLLCSKLWDLLYKLRVLTYYIAPW QITWGSFAFHAFAPFAVPHSAMLVQALLSGLFSTPL NPLLGSVAFIMSYARPLKFWERDYNTKRVDHSNTRLV TQLDRNPGADDNNLNSIFYEHLTRSLQHTLCGDLVLG RWGNYGPGDCF
2358	A	3	301	STATWAGVQWCNLSSLQPLPSGFKPFSCLSLPGSWDH RHLPPCPANFLYCFFLVEMGFHYVGQAGLKLLT/S/G DLCASAPQSAGSTGVNHRVRLGLLIYIP
2359	A	326	1379	PEPHAVQCAELRHQQPRDPQRLQODGSADAPAERKPH CGGERAHGSG\FLAMLLVLGLCGAAYRPTTEEIDLRV GWGNI FQLPFKHVRDYRLRHLVPFFIYSGFEVLFACT

666
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GIALGYGVCSVGLERLAYLLV\AYSLGASAASLLG\L LGLWLPRPVPLVAGAGVHLLLTFF\WAPVPRVLQ HSWILYVAAALWGVGSALNKTGLSTLLGILYEDKERQ DFIFTIYHWWQAVAIFTVYLGSSLHMKAKLE\VLLVT LVAAAVSYLRMEQKLRRGVAPRQPR\IPRQHKVRG\ YRYLQAHNSDES DPEGEHADAAQEEAPPAGPRPGP\E PAGLGRRPCPYEQAQGGD\GP EEG
2360	A	2	1397	LRAGEDMAASASAAAGEEDWVLPSEVEVLES IYLDL QVIKNGRTSPWEIYITLHPATAEDQDSQYVCFTLV QVPAEYPHEVPQISIRNPRGLSDEQIHTILQVLGHVA KAGLGTAMLYELIEKGKEILTDNNIPHGQCVICLYGF QEKEAFTKTPCYHYFHCHCLARYIQHMEQELKAQQQE QEQRQHATTKQKAVGVQCPVCREPLVYDLASLKAAP EPQQPMELYQPSAESLRQQEERKRLYQRQQERGGIID LEAERNRYFISLQQPPAPAEPESAVDVSKGSQPPSTL AAELSTSPAVQSTLPPPLPVATQHICEKIPGTRSNQQ RLGETQKAMLDPPKPSRGPWRQPERRHPKGGECHAPK GTRDTQELPPPEGPLKEPMDLKPEPHSQGVEGPPQEK GPGSWQGP PPRTRDCVRWERSKGRTPGSSYPRLPRG QGAYRPGTRRESLGLESKDGS
2361	A	718	305	SEQEPLLGDTPGSREWDILETEEHYKSRWRSIRILYL TMFLSSVGF SVVMMSIWPYLQKIDPTADTSFLGWVIA SYSLGQMVASPIFGLWSNYRPRKEPLIVSILISVAAN CLYAYLHIPASHNKYYMLVARGLLGIG
2362	A	169	879	MTAEFLSLLCLGLCLGYEDEKKNEKPPKPSLHAWPSS VVEAESNVT LKCAHSQNVTFVLRKVND SGYKQE QSS AENEAEFPFTDLKPKDAGRYFCAYKT TASHEWSESSE HLQLVVTDKHDELEAPSMKTDTRTIFVAIFSCISILL LFLSVFIIYRCSQHSSSSEESTKRTSHSKLPEQEAAE ADLSNMERVSLSTADPQGVTYAELSTSALSEAASDTT QEPPGSHEYAALKV*
2363	A	169	879	MTAEFLSLLCLGLCLGYEDEKKNEKPPKPSLHAWPSS VVEAESNVT LKCAHSQNVTFVLRKVND SGYKQE QSS AENEAEFPFTDLKPKDAGRYFCAYKT TASHEWSESSE HLQLVVTDKHDELEAPSMKTDTRTIFVAIFSCISILL LFLSVFIIYRCSQHSSSSEESTKRTSHSKLPEQEAAE ADLSNMERVSLSTADPQGVTYAELSTSALSEAASDTT QEPPGSHEYAALKV*
2364	A	43	369	AAAWGLAAWGEGPTDATSCWEVGAGGPGNSRPNQTVS MDLNSASTVVLQVLTQATSQDTAVLKPAEEQLKQWET QPGFYSVLLNIFTNHTLDINVRWLAVLYFKHGIDR
2365	A	4272	1534	CHGLQHLTPFRELNLSLQG*EPH*AA*QAVRSEEKSI C*GSPSCHLV LGLVLPVARQSSHSAGPAQSAFR*TGT GSGTPKAAEQSGYWEAYTLGHQHNMFPIQRPLVMK GRRIMCGKCEKG*VSDSVTGGRAVAGEQASQRRTVFT AGGGECLGAKSVRASVFTGNQPGVMGLLNGKRGGCFE SGYLF GFIVIGKIQSLEAKVPLPVNGQTGERASPGNC RIHIVDAVC*SEHH*DHFLAAAFLENSTIIS*VAPGS WQDHAVLQKEVQASVRCRGFESVDTAPAGFWAHSPPG LQGEPTTTSVSLFVLAPQDGE GVPFVEGQLVTVLGLV

667
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VPQSIRHTFVHHTQLFLHPTI*KLGA LDVAF LHL LTLV CSSFN VAYG*GKN GGT LHLQLFAEVNAVTRGSAVQRR PSITISSIHVDTKIQQELHDVMVAGADGVVQWGD PFV VGLAGIFHLIDDPLHQIELSFQRRV*EQCQGVKPD SQ PVPRPLRVGLLQVGPLVRGGGRRVAGRGKRCWRDLLF PWRWGLSHRTRDLLRGGDRGHVVIVLCRLGSLVGGL GTDELLWFGGR*LI I IGI**RGRLSGEWGCGLGRGEL FQVSIGIGVSI VHIGQGDHEVLGGAGLVERGALHATG QGVEALVQQLLDVGPAGALGLCDGAALFQGPGRVQQL PAEGLQVCITLVAQWRMHDGRELGGA EWPWQALHGAA ICGVGGA ILLKALSQYFLKGG*RLWCARGQ*PVKKRQ RRWRG*TRR*NGLTIHCFN*LI*GAVCCRLVILRWCG LLEVHGVYGT*IHCLGSFPGRLWP*PFISQERPNGHC QWEFRLAVPSWKCRWSRWRVRGTWRYGNPLLNLL*GA WLGGAACGGQGGPLSTWQACTGPGQA AFLPPFQGAC RPRTQRCRTWVCPIAWRQLLAYTRD
2366	A	193	366	MYGMLEWPI SMYFVAF LHCFLCSGGNLGDSFQALPEL CANCSSSPRVLCCVVMSP LP*
2367	A	1038	1402	YYQISSLPSTVGN GIFLWLLICIFLAKQGG SRL*FQP FGRPRGGGHLRSGVLGQPGQHGETP/SFFYNSKISPA LWGPPVI PSALGGEAGKSL*PRRQRFQRGGIAPLPSR VRGRAKLFLKKK
2368	A	480	226	MHFLATFALFFIFGVFFLFAVL TNLLLAEEVNIRGGN FLGSFLVHTLFLDQVPGEITHDSHLVLAITINTASPK FSSSIFFYQL*
2369	A	259	941	PVSWSLNSCRFFFFF*DQSLPSVV/QAGSGQ*RN LDS L\QPLASRFK*FSSSRL L\SSW\DYRHMATMARLIFI FLVEMGF\TMLARLV LNFLTSSDPPTSAPFKWLGLQG VKPNTRAVGFN**LGYYSI ILYHSNSPGTDLVFILFI YLFTYLFLRQEONSAAQARVQ*WHNLGSLQSPPPGV\ H*FLCLSLPSSWDYRCAPPHQANFFIFSRDGVSPCWP GWS*TPDLR
2370	A	1676	1197	MALRHLALLAGLLVGVASKSMEN TAQLPECCVDVVG NASCPGASLCGPGCYRRWNADGSASCVRGNGTLPAY NGSECKSFAGPGAPFPMNRSSGTPGRPHPGAPRVAAS LFLGTFFISSGLILSVAGFFYLKRSSKLPRACYRRNK APALQPGERLQ*
2371	A	1078	594	VGMELPAVN LKVI LLGHWLLTTWGCI VFSGSYAWANF TILALGVWAVAQRDSIDAISMFLGGLLATIFLDIVHI SIFYPRVSLTDTGRFGVGMAIL SLL LKPLSCCFVYHM YRERGGELLVHTGFLGSSQDRSAYQTIDSAEAPADPF AVPEGRSQDARGY
2372	A	3	517	HEGRELETGQGRQSSVGAAQGTGVRAGVRAGTTQSGR RRARVSGRLAEVSMASVAWAVLKVLLLLPTQTWSPVG AGNPDCDAPLASALPRSSFSSSELSSSHGPGFSRL NRRDGAGGWTPLVSNKYQWLQIDLGERMEVTAVATQG GYGSSDWVTSYLLMFSDGGRNWK
2373	A	3	517	HEGRELETGQGRQSSVGAAQGTGVRAGVRAGTTQSGR RRARVSGRLAEVSMASVAWAVLKVLLLLPTQTWSPVG AGNPDCDAPLASALPRSSFSSSELSSSHGPGFSRL

668
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				NRRDGAGGWTPLVSNKYQWLQIDLGERMEVTAVATQG GYGSSDWVTSYLLMFSDGGRNWK
2374	A	2	1078	GRVGWELWCMYISPPKDWWDAGDPSLPPIRTPAMIGCS FVVMNRKFFGEIGLLDPGMDVYGGENIELGIKVLCCG SMEVLPCSRVAHIERKKKPYNSNIGFYTKRNALRVAE VWMDYKSHVYIAWNLPLENPGIDIGDVSERRALRKS LKCKNFQWYLDHVYPEMRRYNNTVAYGELRNNAKDV CLDQGPLENHTAILYPCHGWGPQLARYTKEGFLHLGA LGTTLTLLPDTRCLVDNSKSRLPQLLDCDKVKSSLYKR WNFIQNGAIMNKGTRCLENRGLAGIDLILRSCTG QRWTIKNSIK*REGAGALEPGPQDMAAPPNIWTSCPG GETARGRQVLDGPPRASPGQHRDPG
2375	A	2	630	ESNSRCRKMPGERCRGGPARLSLLLDLPTRPLPHPRQ VIDFGSASIFSEVRYVKEPYIQSRFYRAPEILLGLPF CEKVDVWSLGCVMDELHLGWPLYPGNNEYDQVRYICE TQGLPKPHLLHAACKAHFFKRNPHPDAAANPWQLKSS ADYLAETKVRPLERRKYMLKSLDQIETVNGGSVASRL TFPDREALAEHADLKSMVELISAC
2376	A	77	273	PRTGMGCCLPADPAEIRSSPSPSWSTAGSQCWMTS FSPCSCAPCCSSGCACTTGFSVSREKESV
2377	A	1164	464	APWPLPLLRSPQSRPHSLGSLFPPLPGLAELDLQRTL SLQAPPVKEGFLFIHRTKGKGPLMSSSFKKLYFSLTT EALSFAKTPSSKCVNELNQWLSALRKVSINNTGLLGS YHPGVFRGDKWSCCHQKEKTGQCDKTRSRVTLQEWN DPLDHDLEAQLIYRHLLGVEAMLWERHRELSGGAEAG TVPTSPGKVPEDSLARLLRVLQDLREAHSSSPAGSPP SEPNCLLELQT
2378	A	706	951	MRCGWGPLGCLGTGAPAGWMVLGSPRSQLOARWSRA SLSAFGWEIRLRPEGPKAPRQLLLVALESETLGVHGG ATPLHCL*
2379	A	2	456	CVNTFGSYICKCHKGFDLMYIGGKYQCHDIDECSLGQ YQCSSFARCYNVRGSYKCKCKEGYQGDGLTCVYIPKV MIEPSGPIHVPKNGTILKGDGTGNNNWIPDVGSTWWP PKTPYIPPIITNRPTSKPTTRPTPKPTPIPTPPPPR IPP
2380	A	3	1435	LRRHFFFPPSFPPLLPLSLPLSSPLSSFPSPRSAGACW GERLVLQALALRGRPAGSWRGEEAGTAMAPQKHGGGG GGSGPSAGSGGGGFGGSAAVAAATASGGKSGGGSCG GGGSYSASSSSSAAAAAGAAVLPVKKPKMEHVQADHE LFLQAFKPTQIYRFL*TRNLIAPIFLHRTLTYMSHR NSRTNIKRKTFKVDDMLSKVEKMKGEQESHLSAHLQ LTFTGFFHKNDKPSNSENEQNSVTLEVLLVKVCHKK RKDVSCPIRQVPTGKKQVPLNPDLNQTGKGNFPSLAV SSNEFEPNSNHMVKSYSLLFRVTRPGRREFNGMINGE TNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNR RLQLLDGEYEVAMQEMEECPI SKKRATWETILDGKRL PPFETFSQGPTLQFTLRWTGETNDKSTAPIAKPLATR NSES LHQENKPGSVKPTQTIAVKESLTDDLQKK
2381	A	20	1748	KPFNVGLSLNKTERLQLSHGGCKARTAVRAGVFYRAV LQPLTLAQGGLPGSGK/EGSSGCAGTDVGEQASGHR

669
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				ALS*QAVTPAPS*MGHPLSGS*GHQLEPQAGTSPNFA LVTLGHSRPQFPQL*GEALGRRGWQPVS**PGVSIR ET*EAARRGSASARQGRPSS*QGTC*I*RT/AGVKKT PAGQAREGQL*GGTAACGAVGPERVGISPS\QEHGPG GRRGVRVDKDTPAESHPSIPSNKGTSPSRKPAVFPGA PVPPSLTPLSHKATLPSSLTGGRGGGGGKADCSEPG CPVLCQQMPPFHLPLAPASDHFGSAPGLQPPQORKPEG LPGRCSRSDPSGVPTAPESGPGGPEPRP/GTQDALVWP CLGPCSGPSQDLGSGGTCSLCSRHHPLPRPT*VAS S*GQAGLSFAHPSPP/SRAELGQDANATPPSA*R/GS PAQRGINNWGGPVGGAGWAR/PGQEATPAGTEYG*DC PSVGSPQAQDGGQGRRCGGG\PGPW*HH*AHSPCGA AGCWPRCRRSSAADQRAAQGAPPCAGTGAARRARVRC PAGAAGSAAARTNRNRPAG*QSAPPGRTRGS
2382	A	84	428	MSERVERNWSTGGWLLALCLAWLWTHLTALAALQPPTA TVLVQQGTCEVIAAHRCCNRNRIEERSQTVKCSCFSG QVAGTTRAKPSCVDDLLLAHCARRDPRAALRLLLPQ PPSS
2383	A	84	428	MSERVERNWSTGGWLLALCLAWLWTHLTALAALQPPTA TVLVQQGTCEVIAAHRCCNRNRIEERSQTVKCSCFSG QVAGTTRAKPSCVDDLLLAHCARRDPRAALRLLLPQ PPSS
2384	A	1919	3044	HQGPSTPPSWAMSGPPTPLSREDWHQGPSTPPSWAMS EPPT/SSIQGLASGAVHTILLGDVRATYTSIQGVTS VSQVSRAAQMAVPSSRILQLSKPKAPATLLE\EWDPV PKPKPHVSDHNRLHLAKVPRKEGSGKKVGAFPEIKG PEAFRDKARAMESQSNMPFDELLALYGYEASDPISD RESEGGDVPNLPDMTLDKEQIAKDLLSGEEEEETQS SADDLTSPVTSHEASDLFPNRSGCLLAGEAESRGLL PRAQPVPRGAGLADNSRGALLRAHGTVRVGTATATVKP ADAPPESPRDRSRNDSHRPTGPSESERQPQSNQPTL LLRGHGTIRVRTTATVKPADAPAESPRDRSRNDSHG QSSRRSC
2385	A	1206	2266	RHLLTIFHKLKIYKTINKIDFKKKRVTQLLVFCLFLC LFFSSEMVKNQTMVTEFLLLGFLLGPRIQMLLFGFLS LFYVFTLLGNGTILGLISLDSRLHTPMYFFLSHLAVV NIAYACNTVPQMLVNLLHPAKPISFAGCMT*TFLFLS FAHTECLLLVLMsydryVAICHPLRYFIIMTWKVCIT LAITSWTCGSLLAMVHVSILRLPFCGPREINHFFCE ILSVLRLACADTWLNQVVI FAACMFILVGPLCLVLVS YSHILAILRIQS GEGRRKAFSTCSSHL CVVGLFFGS AIVMYMAPKSRHP EEQQKVLFLFYSSFNPM LNPLIYN LRNVEVKGALRRALCKESH
2386	A	1206	2266	RHLLTIFHKLKIYKTINKIDFKKKRVTQLLVFCLFLC LFFSSEMVKNQTMVTEFLLLGFLLGPRIQMLLFGFLS LFYVFTLLGNGTILGLISLDSRLHTPMYFFLSHLAVV NIAYACNTVPQMLVNLLHPAKPISFAGCMT*TFLFLS FAHTECLLLVLMsydryVAICHPLRYFIIMTWKVCIT LAITSWTCGSLLAMVHVSILRLPFCGPREINHFFCE ILSVLRLACADTWLNQVVI FAACMFILVGPLCLVLVS

670
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				YSHILAAILRIQSGEGRRKAFSTCSSHLCVVGLFFGS AIVMYMAPKSRHPPEEQQKVLFLFYSSFNPLNPLIYN LRNVEVKGALRRALCKESH
2387	A	176	371	HFYFCFSDINLAAEPKVNRGKAGVKRSAAEMYGSVTE HPSPSPLLRSGTLLFITALCPSVGIFSF
2388	A	3870	3673	NTQCIPEGLESYYAEQDSSAREKFYTVINHYNLAKQS ITRSVSPWMSVLSEEKLSEQETEAEEKSA
2389	A	1	542	SGSSHASDGSGFQELRICSEDQTPLIAGMCSLPMARY YIIKYADQKALYTRDGQLLVGDPVADNCCAEEKICTLP NRGLDRTKVPIFLGIQGGSRCLACVETEEGPSLQLED VNIEELYKGGEATRFTEFFQSSSGSAFRLEAAAWPGW FLCGPAEPQQPVQLTKESEPSARTKFYFEQSW
2390	A	3	569	ILNERLANYLQKVRMLERENAELESKIQEESENKELPV LCPDYLSTYTTIEELQOKILCTKAENSRLVSQIDNTK LTADDLRAKYEAESVLRQLVESDANGLKQILNVLTLG KADLEAQVQSLKEELLCLKNNHKEEINSLQCQLGERL DIEVTAAPSADLNQVLQEMRCQYEPIMETNRKDVEQW FNTQ
2391	A	3	581	GRRLRSEPRPARPPIARAWPPAPGADGRARRTRVPAP CLPRAPCYGVRPAWRPRPARLRGGLVRWLLSGGPQP RRPRATERPSAGTGAAPRTEPRGRRCRCGRGRG*GP RAWGLALCSPHSCSGAAWGPTTGSQRSWPAVARSWQG DSSRCPALRTTTVTAGSKAALPESAAEVSPMSSSPGR KRSF
2392	C	175	454	MGSLCFLPSLQYWCDELKVEXKTQGRGFPLPGSPASA SHASWTALVKGVGSGQAQEAEGSEEQEIGESPGQSQG VAGAGLGLNEGQVPRMXTR
2393	A	157	396	GGGWTSCSVRFLEQQNQVLETKWELLQQLDLNNCKNN LEPILEGYISNLRKQLETLSGDRVRLDSELRSVRDVV EDYKKR
2394	A	126	561	WKMKKMCNWLRIINYTPDMARAAVDEAIEGLEVWSK VTPLKFTKISKGIADIMIAFRTRVHGRCPRYFDGPLG VLGHAFPPGPGLGGDTHFDEDENWTKDGADLHDNSPF YGHGCLAHAFPPGPGLGGDVHFDNDETRTKDFR
2395	A	126	561	WKMKKMCNWLRIINYTPDMARAAVDEAIEGLEVWSK VTPLKFTKISKGIADIMIAFRTRVHGRCPRYFDGPLG VLGHAFPPGPGLGGDTHFDEDENWTKDGADLHDNSPF YGHGCLAHAFPPGPGLGGDVHFDNDETRTKDFR
2396	A	1	1452	MAELRPSGAPGPTAPPAPGPTAPPAPAFASLFPPGLHAI YGECCRRLYPDQPNPLQVTAIVKYWLGGPDPLDYVSMY RNVGSPSANIPEHWHYISFGLSDLYGDNRVHEFTGTD GPSGFGFELTFRLKRETGESAPPTWPAELMQGLARYV FQSENTFCSGDHVSWHSPLDNSESRIQHMLLTEDPQM QPVQTPFGVVTFQLIVGVCTEELHSAQQWNGQGILEL LRTVPIAGGPWLITDMRRGETIFEIDPHLQERVDKGI ETDGSNLSGVSAKCAWDDLSRPPEDDEDSRSICIGTQ PRRLSGKDTEQIRETLRRGLEINSKPVLPPIINPQRQN GLAHDRAPSRKDSLESDSSTAIIPHELIRTRQLESVH LKFNQESGALIPCLGRLLHGRHFTYKSITGDMAIT FVSTGVEGAFATEEHYPYAAHGPWLQILLTEEFVEKML

671
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				EDLEDLTSPEEFKLPKEYSWPEKKLKVSIPLDVFVDS PLH
2397	A	126	434	MCTKTIPVLWGCFLWNLYVSSSQTIYPGIKARITQR ALDYGVQAGMKMIEQMLKEKKLPDLGSESLEFLKVD YVNYNFSNIKISAFSFPNTSLAFVPGVGI
2398	A	1489	290	FRPLATEPRGSSPVQLVSSTMSVRTLPLLFLNLGGEM LYILDQRLRAQNI PGDKARKVLNDIISTMFNRKFMEE LFKPQELYSKKALRTVYERLAHASIMKLNQASMDKLY DLMTMAFKYQVLLCPRPKDVLLVTFNHLDTIKGFIRD SPTILQQVDETLRQLTEIYGGLSAGEFQLIRQTLLIF FQDLHIRVSMFLKDKVQNNNGRFVLPVSGPVPWGTEV PGLIRMFNNKGEEVKRIEFKHGGNYVPAPKEGSFEFY GDRVCLKGTNMYSVNQPVETHVSGSSKNLASWTQESI APNPLAKEELNFLARLMGGMEIKKPSGPEPGFRLNLF TTDEEEEQAALTRPEELSYEVINIQTQDQQRSEELA RIMGEFEITEQPRRLSTSKGDDLLAMMDEL
2399	A	1489	290	FRPLATEPRGSSPVQLVSSTMSVRTLPLLFLNLGGEM LYILDQRLRAQNI PGDKARKVLNDIISTMFNRKFMEE LFKPQELYSKKALRTVYERLAHASIMKLNQASMDKLY DLMTMAFKYQVLLCPRPKDVLLVTFNHLDTIKGFIRD SPTILQQVDETLRQLTEIYGGLSAGEFQLIRQTLLIF FQDLHIRVSMFLKDKVQNNNGRFVLPVSGPVPWGTEV PGLIRMFNNKGEEVKRIEFKHGGNYVPAPKEGSFEFY GDRVCLKGTNMYSVNQPVETHVSGSSKNLASWTQESI APNPLAKEELNFLARLMGGMEIKKPSGPEPGFRLNLF TTDEEEEQAALTRPEELSYEVINIQTQDQQRSEELA RIMGEFEITEQPRRLSTSKGDDLLAMMDEL
2400	A	1214	1357	NKINMFIAALFTIAKT\WNQPK\CPTMIDWIKKRGSS RVASSSSPTRTR
2401	A	85	396	MILINFREICLKVLHTPLCVSGGCVLLYILALTCCYT NSLLISHLPPLSLPTETQTHLFMYRVLKVRKDIKNHV FHPTYLVAKETETYGEELIPLPPCREHQD*
2402	A	919	1439	KLKDDFFEMEYCSVAQAGVQWSLQPPSPWFKQFSYVS LPSSWDYSHLPPCPANLFLVEMRFHLVGQAGLKLLTS GDPPASASRSAGIIGVSHHAWPKIKRFYETKWLPILS IQLLSGLFIWALLFFCFVLHFCIIWGNLSLEVFPESV CRHNKICVLCTQKHNVSYESITQPV
2403	A	74	226	MSSWPRMLAHCFYLLKALSSSYLIKEMTIMPGTLLST LCILTHLNLPTPL*
2404	A	255	369	PTESAPGLGFCFPDFGQSLPNEKQTSAL\LSDHQQSQ LC
2405	A	5671	1873	GREREEELQWRRRRRQRRGAAAPAAPAGGIEAVNMAS ASYHISNLLKMTSSDKDFRFRMATNDLMTELQKDSIK LDDDSERKVVKMILKLLEDKNGEVQNLAVKCLGPLVS KVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTIV GELPPASSGSALAANVCKKITGRLTSAIAKQEDVSVQ LEALDIMADMLSRQGGLLVNFHPSILTCLLPQLTSPR LAVRKRTIIALGHLVMSCGNIVFDLIEHLLSELSKN DSMSTTRYIQCIAAISRQAGHRIGEYLEKIIPLVVK FCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIIN

672
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				ICLKYLTYDPNPNYDDEDEDENAMDADGGDDDDQGSDD DEYSDDDDMSWKVRRAAAKCLDAVVSTRHEMLPEFYK TVSP\ALISRFKEREENVRADV FHAYLSLLKQTRPVQ SWLCDPDAMEQGETPLTMLQSQVPNIVKALHKQMKEK SVKTRQCCFNMLTELNVNVLPGALTQHIPVLVPGIIFS LNDKSSSSNLKIDALSCLYVILCNHSPQVFHHPVQAL VPPVACVGDPFYKITSEALLVTQQLVKVIRPLDQPS SFDATPYIKDLFTCTIKRLKAADIDQEVKERAI SCMG QIICNLGDNLGSDDLNTLQIFLERLKNELTRLTTVKA LTLIAGSPLKIDLRPVLGEGVPILASFLRKNQRALKL GTL SALDILIKNYSDSLTAAMIDAVLDELPLISESD MHVSQMAISFLTTLAKVYPSSLSKISGSILNELIGLV RSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRML TGPVYSQSTALTHKQSYYSIAKCVAALTRACPKEGPA VVGQFIQDVKNRSTDSIRLLALLSLGEVGHHDLSG QLELKS VILEAFSSPSEEVKSAASYALGSI SVGNLPE YLPFVLQEITSQPKRQYLLHSLKEI ISSASVGLKP YVENIWALLLKHCECAEEGTRNVVAECLGKLTIDPE TLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPID PLLKNCIGDFLKTLED PDLNVRVALVTFNSAAHNKP SLIRDLLD TVLPHLYNETKVRKELIREVEMGPFKHTV DDGLDIRKAAFECMYTLLDSCLDRLDIFEFNLHVEDG LKDHYDIKMLTFLMLVRLSTLCPSAVLQRLDRLVEPL RATCTTKVKANSVKQEFQDELKRSAMRAVAALLTI PEAEKSPLMSEFQSQISSNP ELAAIFESI QKDSSTN LESMDTS
2406	A	1	824	THACALISSRFIILSSFHVILNKTKHTCIHTHSLTLK MQDEERYMTLNVQSKKRSSAQTSQ LTFKDYSVTLHWY KILLGISGTVNGILTTLTILSLILLVSQGVLLKCQKGS CSNATQYEDTGD LKVNNGTRRNISNKDLCASRSADQT VLCQSEWLKYQKCYWFSNEMKSWSDSYVYCLERKSH LLIIHDQLEMSLV\QF*AFIQKNLRQLNYVWIGLNFT SLKMTWTWVDGSPIDSKI FFIKGPAKENS CAAIKESK IFSETCSSVFKWICQY
2407	A	182	418	MCCELLAVVIATLI IKIGLVVLLYFIKLLIHIEFIKR HSILKCESIFNLNVGIRMYPGQVNF CETLQMLDGFGR IFQTK
2408	A	65	320	LQMSSLPTAAPALD VDWQSSTTFASCSTDMCIHVCRL GCDRPVKTFQGHTVSESSCHWSRVCENVMWEPILVCL ELKATAAADQL
2409	A	923	358	ALSCGPFPQPLGDKLFRWWLLPLSRFLMRVLDSYGDD YRASQFTIVLEVSVGPPGGSGTGSSGPTHHLPPPPAC QDEGSQGTDAPTPGNAENEPPEKETLSPPRRTPAPPE \PGSP\APGEGPSGRKRRRVPRDGRPAGNALTPELAP VQIKVEEDFGFEADEALDSSWVSRGPDKLLPYPTLAS PAFD
2410	A	923	358	ALSCGPFPQPLGDKLFRWWLLPLSRFLMRVLDSYGDD YRASQFTIVLEVSVGPPGGSGTGSSGPTHHLPPPPAC QDEGSQGTDAPTPGNAENEPPEKETLSPPRRTPAPPE \PGSP\APGEGPSGRKRRRVPRDGRPAGNALTPELAP

673
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				VQIKVEEDFGFEADEALDSSWVSRGPDKLLPYPTLAS PAFD
2411	A	923	358	ALSCGPPFPQPLGDKLFRWWLLPLSRFLMRVLDSYGDD YRASQFTIVLEVSVGPPGGSGTGSSGPTHHLPPPPAC QDEGSQGTDAPTPGNAENEPPEKETLSPPRRTPAPPE PGSP\APGEGPSGRKRRRVPRDGRPAGNALTPELAP VQIKVEEDFGFEADEALDSSWVSRGPDKLLPYPTLAS PAFD
2412	A	12	1154	GILRQKEREERNRIHKKEILFLEHLLVVPSEMSSLG KVQTVLGLVEPSKLGRITLTHEHLAMTFDCCYCPPPPC QEAIKKEPIVMKNLYWIKNAYSHKENLQLNQETAI KEELLYFKANGGGALVENTTTGISRDTQTLKRLAET GVHIIISGAGFYVDATHSSETRAMSVEQLTDVLMNEIL HGADGTSIKCGIIGEIGCSWPLTESERKVLQATAHAQ AQLGCPVIIHPGRSSRAPFQIIRILQEAGADISKTVM SHLDRTILDKKELLEFAQLGCYLEYDLFGTELLHYQL GPDIDMPDDNKRIRRVRLVVEEGCEDRILVAHDIHTK TRLMKYGGHGYSHILTNVVPKMLLRGITENVLDKILI ENPKQWLTFK
2413	A	575	759	SVYSASSCKCCNYRKTEQIPDCEQPPASSMPERPSHE SQTPQMMPLSAPSRAEELGQRPG
2414	A	131	1677	VRGDDLTRALRARRRRSGSGSNFRVVEPQATGILLFL PPPPVCPAPLPLSLLFPAPPKMNSSDEEKQLQLITS LKEQAIGEYEDLRAENQKTKEKCDKIRQERDEAVKKL EEFQKISHMVIEEVNFMQNHLEIEKTCRESAEALATK LNKENKTLKRISMLYMAKLGPDVITEEINIDDEDSTT DTDGAAETCVSVQCQKQIKELRDQIVSVQEEKKILAI ELENLKSCLVEVIEEVNKKVQEKTVLNSEVLEQRKVL EKCNRVSMMLAVEEYEMQVNLELEKDLRKKAESFAQE MFIEQNKLKRQSHLLQSSI PDQQLLKALDENAKLTQ QLEEEERIQHQQKVKELEEQLNETLHKEIHNKQOLE LLEEDKKELELKYQNSEEKARNLKHSVDELQKRVNQS ENSVPPPPPPPPPLPPPPPNPIRSLMSMIRKRSHPSG SGAKKEKATQPETTEEVTDLKRQAVEEMMDRIKKGVH LRPVNQTARPKTKPESSKGCESAVDELKGILASQ
2415	A	1157	918	RSGVPDQPGQHGEAPSLKIQNLAGRSGGPL*SQLLR RENRLNLGGGLP*AKIAPRLHPCTPAWVTDRDSVSKK KILFP
2416	A	70	222	MFCSFPLILQVYPTWKNPNWHLTFHTSVFSFPKGVR SLARGIPDHLHSA*
2417	A	163	531	MQQMMWAGLLCPQLEWLQGRACRPCGLLASDAAALWF RGGISAWEDSCAVSNIRHEAYNCHLSVFLNRCANELT VQFLIILAFQIMLSCAVIAPAVPVFQRLTLKRSGRTS LGSTGRLHFCK*
2418	A	60	266	MKRLRFVLRVFMQTAFTGAHTITNYSRRRLYISPLS HFFMNSGSSAQSVLSHSYVSQIFFKNVSKYF*
2419	A	218	1885	QSDLSTRTQLARLLFCAKTGELVGTMKIFCSRANPTT GSVEWLEEDEHYDYHQEIARSSYADMLHDKDRNVKYY QGIRAAVSRVKDRGQKALVLDIGTGTGLLSMMAVTAG ADFCYAIEVFVKPMADAAVKIVEKNGFSDKIKVINKHS

674
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				TEVTVGPEGDMPCRANILVTELFDTTELIGEGALPSYE HAHRHLVEENCEAVPHRATVYAQLVESGRMWSWNKLF PIHVQTSLGQVIVPPVDVESC PGAPSVCDIQLNQVS PADFTVLSVDLPMFSIDFSKQVSSSAACHSRRFEPLT SGRAQVVLSSWWDIEMDPEGKIKCTMAPFWAHS DP EEM QWRDHWMQCVYFLPQEEPVVQGSALYLVAHHDDYCVW YSLQRTSPEKNERVRQMRPVCDCQAHLLWNRPRFGEI NDQDRTDYVQALRTVLKPDSVCLCVSDGSLLSVLAH HLGVEQVFTVESSAASHKLLRKIFKANHLEDKINIE KRPELLTNEDLQGRKVSLLLGEPPFTTSLLPWHNLYF WYVRTAVDQHLGPGAMVMPYAASLHAVVVEFKDLWRI R
2420	A	2121	1148	HYLGSLELGQCGQLSPLPCGLQVALYKSVPTRLLSRA WGRLNQVELPHWLRRPVYSLYIWTFGVNMKEAAVEDL HHYRNLSEFFRRKLKPQARPVCGLH SVI SP SDGRILN FGQVKNCEVEQVKGV TYSLESFLGPRMCTEDLPFPPA ASCDSFKNQLVTREGNELYHCVIY LAPGDYHCFHSPT DWTVSHRRHFPGSLMSVNP GMARWIKELFCHNERVVL TGDWKHGFFSLTAVGATNVGSIRIYFDRDLHTNSPRH SKGSYNDFS FVTHTNREGVPMRKGEHLGEFNLGSTIV LIFEAPKDFNFQLKTGQKIRFGEALGSL
2421	A	195	859	GCPGCCSPRCCLAGAHSDGPGPGSSCSSRGRQVSGNR AWTGPSSQARRSPGLRGQRLAGARPPSWPE/EDSRV PGKDKL*GKELEISA*SQPPSARPPSGCTAPGANRNS WTNSSERILRAHF/APLPSPPPPLEAGG/LPP*GAT RGPSAVPSFPSVSGDWGGPVEAGRAGSRAEGEPGRAL APSLLCSLPPRFAGSQALGLPWAVTAERWQELRASEL RNR
2422	A	87	594	KCLRKSDEALNRVLQQT\RVP PKMRGTS LHSRRGKP EAPKGS PQINRKSGQEMTAVMQSGRPRSSSTTDAPTG SAMMEIACAAAAAAACLPGEEGTAERIERLEVSSLA QTSSAVASSTDGSIHTDSVDGTPDPQRTKAAIAHLQQ KILKLTEQIKIAQTARRNRRPG
2423	A	2230	990	NSSGVKLLQALGLSPGNGKDHSLHSRNDLEEAFIHF MGKGAAAEFFSDKETFDIAQVASEFPGAQH YVGGN AALIGQKFAANS DLKVL LCGPVGPKLHELDDNVFVP PESLQEVDEFHLILEYQAGEEWGQLKAPHANRFIFSH DLSNGAMNMLEV FVSSLEEFQPD LGGLSGLHMMEGQS KELQRKRLLEVVT S I SDIPTGIPV\HLELG\ SMTNRE LMSSIV\LQQVFPAVTS LGLNEQELLFLTQSASGPHS SLSSWNGVPDVG MVSDILFWILKEHGRSKSRASDLTR IHFHTLVYHILATVDGHWANQLA A VAAGARVAGTQAC ATETIDTSRVSLRAPQEFMTSHSEAGSRIVLNPNKPV VEWHREGISFHFTPV LVCKDPIRTVGLGDAISAEGLF YSEVHPHY
2424	A	122	505	MLWELVLLGEPLVVMAPSPSESSETVLALVNCISPLK YFSDFRPYFTIHDSEFKEYTTRTQAPPSVILGVTNPF FAKTLQHWPHIIRIGDLKPTGEIPKQVKVKKLKNLKT LDSKPGVYTSYKPYSN*
2425	A	2	271	GSVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLV

675
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SQLIRAGKPYQLQVALPPVSPQIYPNERHSIRCPESG EHYEVTLHFLQEYL
2426	A	2	271	GSVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLV SQLIRAGKPYQLQVALPPVSPQIYPNERHSIRCPESG EHYEVTLHFLQEYL
2427	A	2	271	GSVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLV SQLIRAGKPYQLQVALPPVSPQIYPNERHSIRCPESG EHYEVTLHFLQEYL
2428	A	245	392	GPGCIPAALLQPPKDDKKKKDAGKSAKKDKDPVNKSG GKAKKKVEIRPL
2429	A	138	1671	EAVQVLIKHSADVNRDKNWQTPLVHAAANKAVKCAE VIIPLLSSVNVSDRGGRTALHHAALNGHVEMVNLLLA KGANINAFDKKDRRALHWAAYMGHLDVVALLINHGAE VTCKDKKGYTPLHAAASNGQINVVKHLLNLGVEIDEI NVYGNTALHIACYNGQDAVVNELIDYGANVNQPNNG FTPLHFAAASHTGALCLELLVNNGADVNIQSKDGKSP LHMTAVHGRFTRSQTLIQNGGEIDCVDKDGNTPLHVA ARYGHELLINTLITSGADTAKCGIHSMPHLHLAALNA HSDCCRKLLSSGQKYSIVSLFSNEHVLSAGFEIDTPD KFGRTCLHAAAAGGNVECIKLLQSSGADFHKDKCGR TPLHYAAANCHFHCIETLVTTGANVNETDDWGRTALH YAAASDMDRNKTILGNAHDNSEELERARELKEKEATL CLEFLLQNDANPSIRDKEGYNSIHYAAAYGHRQCLEL LLERTNSGFEESDSGATKSPLHLAVSEMP
2430	A	1266	210	PWAVSQLASGG\ATIPGIRGAGRSRPPGILVPACTSE G/P/SSQYNFIADVVEKTAPAVVYIEILDRHPFLGRE VPISNGSGFVVAADGLIVTNAHVVADRRRVRVRLLSG DTYEAVVTAVDPVADIATLRIQTKEPLPTLPLGRSAD VRQGEFVVAMGSPFALQNTITSGIVSSAQRPARDLGL PQTNVEYIQTDAIDFGNSGGPLVNLGDGEVIGVNTMK VTAGISFAIPSDRLREFLHRGEKKNSSSGISGSQRRY IGVMMLTLSPSILAEQLREPSFPDVQHGVLIHKVIL GSPAHRAGLRPGDVILAIGEOMVQNAEDVYEAVRTQS QLAVQIRRGRETTLTYVTPEVTE
2431	A	80	403	MLWFSGVGALAERYCRRSPGITCCVLLLLNCSGVPM LASSFLTGSVAKCENEVEVLQIPFITDNPCIMCVCLN KEVTCKREKCPVLSRDCALAIKQRGACCEQCKGC
2432	A	469	1020	GISGKAGGSMRSGSVCSGAAAMPIEEPALRSWQRPFL KWAGGKYSLLPELDRLLIPAGKRLLIEPFVGGGSVFLNS DKHERFLLADVADLINLYQMLAVVPDSVIYEAMKAF RHLNDAENYTLIREAFNAQRLDAVERAAAFLYLNRHC FNGLIRYNLDGFFQQGH*ER*RQVFPRQSVVQRTDS
2433	A	1	266	GHFRVPALGYLDVRIVDTDYSSFAVLYIYKELEGALS TMVQLYSRTQDVSPQALKAFQDFYPTLGLPEDMMVML PQSNACNPESKEAP
2434	A	2	1318	LRKEGRRCRRGSNRGVWAAPAEGLGGRGMLGVRCLLRS VRFCSSAPFPKHKPSAKLSVRDALGAQNASGERIKIQ GWIRSVRSQKEVLFLHVNDGSSLESQVADSGLDSR ELTFGSSVEVQGQLIKSPSKRQNVELKAEKIKVIGNC DAKDFPIKYKERHPLEYLRQYPHFRCRTNVLGSILRI

676
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				RSEATAAIHSFFKDSGFVHIHTPIITSNDSEGAGELF QLEPSGKLKVPEENFFNVPAFLT VSGQLHLEVMSGAF TQVFTFGPTFRAENSQSRRHLAEFYMIEAEISFVDSL QDLMQVIEELFKATTMMVL SKCPEDVELCHKFIAPGQ KDRL*HMLKNNFLIISYTEAVEILKQASQNF TFTP EW GADLRTEHEKYL VKHCGNIPVFVINYP LTLKPFYMRD NEDGPQELEG SVA*HSLGLMILLSIVVIGQP
2435	A	58	501	GNKAFVCFYLSQLENYGMPFSRTEDGKIYQRAFGGQS LKFGKGRQAHRC CVADRTGHSILHTSYGRSLRYDTS YFVEYFALDLLMENRECRGVIAQCNE DGSIHHRANKN TVVATG*ESNFYFISFVKMNKFLLECLYFKENRGIVE
2436	A	3	717	DSL DNHRCRGDLTKTYSLEAYDNWFNCL SMLVATEVC RVVKKKHRTRMLEFFIDVARECFNIGNFN SMMAIISG MNLSPVARLKKTWSKVKTAKFDVLEHHMDPSSNFCNY RTALQGATQRSQMANS SREKIVIPVFNLFVKDIYFLP QNP\SNHLPNGHINFKKFWEISRQIHEFMTWTQVECP FEKDKKIP\SYLLTAPHPTARKLSSSPSF ESEGPENH MEKDSWKT LRTTLLNRA
2437	A	130	726	ITCCGYDALSSIRKNLCCLWICSKPYSLLMGEGDAFW APSVLPHSTLSTLSHHPQPQFGRGMESKVSQGLNVT LTIRLLMHGKEVGSIIIGKKGETVKKMREESGARINIS EGNCPERIVTITGPTDAIFKAFAMIAYKFEEDI INSM SNSPATSKPPVTLRLVVPASQCGSLIGKGGSKIKEIR EVTGPSQPGPLRSL
2438	A	401	249	DTLIYTCAPEFDFMEKATPLRYTKTLLLPVVMVITCF IFKKTVRDISCVLA
2439	A	1671	429	TGGRVGGSRSRRALPLPAPVEAGVLT SAGPSGVVWQR IEDTTKMAAVSGLVRRPLREVSGLLKRRFHWTAPAAV QV\TVRDAINQGMDEELERDEKVFLLGEEVAQ\YDGA YKVSRLWKKYGDKRII\DTPISEM GFAWELLVGA AI GWGLRPILLNLWTFNFSM\QAI\ DQVINSAAKTYYM\ SG\GLQPVLIVSWGPN\ GASAGVAAQHSQCFAAWYGH CPGLKVVP\WTS*DAKGLIKSAIRDNNPVVALENEL MYGVPF\ EFPPEAQSKDF\ LIPIGKAKIEMHGTHITV VSHSRPVG\HCLRSLPAS/VLSKEGVEC\ EVINMRT\ IRP\MDMET\IEA\SVMKTKFIL* LWEGGW PQFG\VG A\EICARIM\EGPAFNF\LDAPAVRVTGADVMPYAK ILEDNSIPQVKDII FAIKKTLNI
2440	A	66	1349	APNSESQTQGPLPTPANLFWTRRANPDPTTSMSATDR MGPKAVPGLRLALLLLGLGTPKSGVQGGQEG LDFPEY DGVDRVINVNANKYKNVFKKYEVLALLYHEPPEDDKA SQRQFEMEELILELAAQVLEDKGVGFG LVDSEKDAAV AKKLGLTEVDSMYVFKGDEVIEYDGEFSADTIVEFLL DVLEDPVELIEGERELQAFENIEDEIKLIGYFKSKDS EHYKAFEDAAEEFHPYIPFFATFDSKGAKKLT LKLN IDFYEA FMEEPVTIPDKPNSEEEIVNFVEEHRRSTLR KLKPESMYETWEDDMDGIHIVAF AEADPDGFEFLET LKAVAQDNTENPDL SIWIDPDDFPLLVPYWEKTFDI DLSAPQIGVVNVTDADRLWMEMDDEEDLPSAE ELEDW LEDVLEGEINTEDDDDDDDD

677
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
2441	A	1002	2209	VYPYNPLAFFRERLQPNCPFHSSSEGSSGKLS*PPSPS FTSSLCDSTSGFGASSTH*HS*IRATLISSAFTLA VAWAALLCPPISSCSSETWLSLRQMGSEPKVQPSCE ASPSSVHLPPLPSWAVSVQASPGSSPSMGPRGSSVSP PLAGGEAGLPTSGNPPNSSPWASGQGGWASLSLTSL SQLSGWMAAA*LGSFSSSSSSFSWGTWLSPFVSSSITG AESTGTSTDAVSNFLSAFKEPEAVMGSGSSWAGSSSS RVPPNSSSDEHVPGPSAVSSVATGFTTGSSTLEIITC SVPSGGGLGPGRERLSPLANELGTSGCFSSSDSWNTS LLRVSLPGTPGRMAEALLAGLAWFDPVGGFRSVKLDL LSLGKAMLSSNKLCCFKIAASFITFRVSSRI
2442	A	1	933	MGSRLLCWVLLCLLGAGPVKAGVTQTPKHLITATGQQ VTLRCSPRSGDLVSWYQQSLDQGLQFLIQYNGEER AKGNILERFSAQQFPDLHSELNLSSLELGDSALYFCA SSVKVGTGELFFGEGSRLTVLEDLKNVFPPEVAVFEP SEAEISHTQKATLVCLATGFYPDHVELSWVNGKEVH SGVSTDPQPLKEQPALNDSRYCLSSRLRVSATFWQNP RNHFRCQVQFYGLSENDEWTQDRAKPVTQIVSAEAWG RADCGFTSESYQQGVLSATILYEILLGKATLYAVLVS ALVLMAMVKRKDSRG
2443	A	368	18	SRTPENYLKSSIDS AHRQKRRTIPSAKGFPGFFRA AKLLCQSLSPFMTGRAP*ALAGDTSAFMALLPRTHLS ATPAVCPPPETFISSVFVASLFTILELKYHLLREAFP LLPS*N
2444	A	5	235	DSSRMSYQQQCKQPCQPPVCPPTPKCPEPCPPPKCP EPCPPKCPQPCPPQCCQKYPVTPSPPCQSKYPPK SK
2445	A	82	2929	TRTKRRLGREKAMASPPRGWGCCELLLPFMLLGTLC PGSGQIRYSMPPELDKGSFVGNIADLGLPEQELAER GVRIVSRGRTQLFALNPRSGSLVTAGRIDREELCAQS PLCVVNFNILVENKMKIYGVEVEIIDINDNFRFRDE ELKVKVNENAAAGTRLVLPFARDADVGVNSLSYQLS SNLHFSLDVVS GTDGQKYPELVLEQPLDREKETVHDL LLTALDGGDPVLSGTTHIRVTVDANDNAPLFTPSEY SVSVPENIPVGTLLMLTATDPDEGINGKLTYSFRNE EEKISETFQLDSNLGEISTLQSLDYEEESRYLMEVVA QDGGALVASAKVVVTVDVNDNAPEVILTSLTSSISE DCLPGTVIALFSVHDGDSGENGEIACSI PRNL PFKLE KSVDNYYHLLTTRDLREETS DYNITLTVMDHGT PPL STESHIPLKVADVNDNPPNFPQASYSTSVTENNPRGV SIFSVTAHDPDSGDNARVTYSLAEDTFQGAPLSSYVS INSDTGVLIALRSFDYEQLRDLQLWVTASDSGNPPLS SNVSLSLFVLDQNDNTPEILYPALPTDGSTGVELAPR SAEPGYLVTKVAVDKDSGQNAWLSYRLLKASEPGLF AVGLHTGEVRTARALLDRDALKQSLVVAVEDHGPPL SATFTVTVAVADRI PDILADLGSIKTPIDPEDLDLTL YLVVAVAAVSCVFLAFVIVLLVLRRLRRWHKSRLLOAE GSRLAGVPASHFVGVDGVRAFLQTY SHEVSLTADSRK SHLIFPQPNYADTLLSEESCEKSEPLLM SDKVDANKE ERRVQQAPPNTDWRFSQAQRPGTSGSQNGDDTGTWPN

678
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				NQFDTEMLQAMILASASEAADGSSTLGGGAGTMGLSA RYGPQFTLQHVLOGELGSDYRQNVYIPGSNATLTNAAGKRDGKAPAGNGNKKKSGKKEKK
2446	A	61	241	ANLGPTAPPRSGPVLGAGEKGRGEMRRAPFFLSAGGLETPPPSAALLWAPGRRADEISGL
2447	A	1	306	CGCGSCGGCGGRCGGGCGGGCSGGCGGGCGGGCGGGCGGCGSCTTCRCYRVGCCSSCCPCCRGCCGGCCSTPVICCC RRTCGSCGCGYGKGCCQQKCCCQKQCCC
2448	A	3	761	YAKLGTRDPSKLCRHSCLKLECNEVFQDETSLATHFQ QAADTSGQQMKKHPCRQCDKSFSSSHSLCRHNRIKHK GIRKVYACSHCPDSRRTFTKRLMLEKHVQLMHGKIDP DLKE/TDRCHQ*GGNRNKRRH*GPQSQAQEVGRTSSGV QASPRSNHSTTEKAENQ\FFKVHKCAVCGFTTENLLQ FHEHIPQHKSDGSSYQCRECGLCYTSHVSLSRHLFIV HKLKEPQPVSKQNGAGEDNQQENKPSHEGGIP
2449	A	2740	2525	MIETWLWLLLLNVGGTGQWSGPTFRRENVLPAAHIGP KYGPLLPSTAKGTVKVSCPSSTPHPLQKGKTPD*
2450	A	656	513	MSLLLPPLALLLLAALVAPATAATAYRPDWNRLSGL TRARVETCGG*
2451	A	42	266	KLILLKIQYFNLLMKCCFRIKGGLEQRPERVKPFMT GAAEQIKHILANFKNYQVNTLSIWIKGLYNFNCKSKN
2452	A	6	664	LPGRPTRAPTRPAEHSIVGTRLVSCQLQPSQPNADQG KLTTMRIAVICFCLLGITCAIPVKQADSGSSEKQLY NKYPDAVATWLNPDPSQKQNLAPQTLPSKSNESH MDDMDDEDDDDHVDSQDSIDSNDSDVDVDDT\DDSHQS DESHHSDES\D\ELVTDFTDLPATEVFTPVVPTVDT YDGRGDSVYGLRSKSKKFRRPDIQYPDATDEDITS
2453	A	68	348	IQGMHFAAGRLSTKTFCTGHGSPVDICTAKPRDIPMN PMGIYRSPEKKATEDEGSEQKIPEATNRDVEPTKAN SRFATTFYQHLADSKNDND
2454	A	5214	352	MAKSGGCGAGAGVGGNGALTWVNNAAKKEESETANK NDSSKLSVERVYQKKTQLEHILLRPDTYIGSVEPLT QFMWVYDEEDVGMNCREVTFVPGLYKIFDEILVNAADN KQRDKNMTCIKVSIDPESNII SIWNNKGKIPVVEHKV EKVYVPALIFGQLLTSSNYDDDEKKVTGGRNGYGAKL CNIFSTKFTVETACKKEYKHSFKQTMNMNMKTSEAKI KHFDGEDYTCTFQPDLSKFKMEKLDKDIVALMTRRA YDLAGSCRGVKVMFNGKKLPVNGFRSYVDLYVKDKLD ETGVALKVIHELANERWDVCLTLSEKGFQQISFVNSI ATTKGGRHVDYVVDQVVGKLI EVVKKKNKAGVSVKPF QVKNHIWVFINCLINPTFDSQTKENMTLQPKSFGSK CQLSEKFFKAASNCGIVESILNWVKFKAQTQLNKKCS SVKYSKIKGIPKLDDANDAGGKHSLECTLILTEGDSA KSLAVSGLGVIGRDYGVFPLRGKILNVREASHKQIM ENAEINNI IKIVGLQYKKSYYDAQSLKTLRYGKIMIM TDQDQDGSIIKGLLINFIIHNWPSLLKHGFLEEFITP IVKASKNKQELSFYSIPEFDEWKKHIENQKAWKIKYY KGLGTSTAKEAKEYFADMERHRILFRYAGPEDDAIT LAFSKKKIDDRKEWLTNFMEDRRQRLHGLPEQFLYG TATKHLTYNDFINKELILFSNSDNERSI PSLVDGFKP

679
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GQRKVLFTCFKRNDKREVKVAQLAGSVAEMSAYHHGE QALMMTI VNLAQN FVGSNNINLLQPIGQFGTRLHGGK DAASPRYIFTMLSTLARLLFPAVDDNLLKFLYDDNQR VEPEWYIPIIPMVLINGAEGIGTGWACKLPNYDAREI VNNVRRMLDGLDPHPMLPNYKNFKGTIQELGQNQYAV SGEIFVVDNRNTVEITELPVRTWTQVYKEQVLEPMLNG TDKTPALISDYKEYHTDTTVKFVVKMTEEKLAQAEAA GLHKVFKLQTTLTCSNMFVDFHMGCLKKYETVQDILK EFFDLRLSYGLRKEWLVGMLGAEFTKLNNQARFILE KIQGKITI*NRSKDLIQMLVQRGYESDPVKAWKEAQ EKAAEEDETQNQHDDSSSDSGTPSGPDFNYILNMSLW SLTKEKVEELIKQRDAKGREVNDLKRKSPSDLWKEDL AAFVEELDKVESQEREDVLGMSGKAIKGKVGKPKVK KLQLEETMPSPYGRRIIPEITAMKADASKKLLKKKKG DLDTAAVKVEFDEEFSGAPVEGAGEEALTPSPVINKG PKPKREKKEPGTRVRKTPTSSGKPSAKKVKKRNPWSD DESKSESDLEETEPVVI PRDSLRRRAAAERPKYTFDF SEEEEDDDADDDDDNNDLEELKVKASPI TNDGEDEFV PSDGLDKDEYTFSPGKSKATPEKSLHDKKSQDFGNLF SFPSYSQKSEDDSAKFDSNEEDSASVFSFGLKQTD KVPSKTVAACKGKPSSDTVPKPKRAPKQKKVVEAVNS DSDSEFGIPKKTTPKGKGRGAKKRKASGSENEGDYN PGRKTSKTTSKPKKTSFDQSDVDIFPSDFPTEPPS LPRTGRARKEVKYFAESDEEEDDVDFAMFN
2455	A	2	154	FKIQKTRLQREGFDPQTS DRLF FLDL KQGHYLP LNE AVYTRICSGAFAL
2456	A	483	765	FQGQRMAGEQKPSSNLLEQFILLAKGTSGSALTALIS QVLEAPGVYVFGELLELANVQELAEGANAAYLQLLNL FAYGTYPDYIANKESLP ELY
2457	A	9	422	ESRERSGNRRGAEDRGTCGLQSPSAMLGAKPHWLP GP LHSPGLPLVLVLLALGAGWAQEGSEPVLLGECLVVC EPGRAAGGPGGAALGEAPPGRVAFAAVRSHHHEPAG ETGNGTSGAIYFDQVLVNEG GGFDRAS
2458	A	64	435	GRGVCVAAWSQRSIAGNNDYRLFHKMSNSHPLRPFTA VGEIDHVHILSEHICALLIGEEYGDVTFVGEKKRFPA HRVILAARCQYFRALLYGGMRESQPEAEIPLQDTTAE AFTMLLXYIYTGR
2459	A	126	434	MCTKTIPVLWGCFLWNLVSSSQTIYPGIKARITQR ALDYGVQAGMKMIEQMLKEKKLPDLSGSESLEFLKVD YVNYNFSNIKISAFSFPNTSLAFVPGVGI
2460	A	126	434	MCTKTIPVLWGCFLWNLVSSSQTIYPGIKARITQR ALDYGVQAGMKMIEQMLKEKKLPDLSGSESLEFLKVD YVNYNFSNIKISAFSFPNTSLAFVPGVGI
2461	A	126	434	MCTKTIPVLWGCFLWNLVSSSQTIYPGIKARITQR ALDYGVQAGMKMIEQMLKEKKLPDLSGSESLEFLKVD YVNYNFSNIKISAFSFPNTSLAFVPGVGI
2462	A	3	1057	EEEQECRPAIKTSDIDNPSHF EKQYESSSSSTHSDRS SDGEQDFVSSILPGNRPNSTNIKPQLHQKSIMKKKAG HKANSKH*D*EQTVVDVTEQLGDCKLDSQEKDATCEL PLQKVNTQSSSNSTLPGR LKASENSESEYSRSEITLV

680
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				GISKKSAEHFKRKFAKSNQVSRSVSSSVQVCPEVGKR NLLKVLKETLIEWKTEETLRFLYGQNYASVCLKPEAS LVKEELDEDDIISDPDSHFPAWRESQNSLDESLPFRG SGTAIKPLPSYENLKKETEKLNLRIREFYRGRYVLGE ETTKSQDSEEHDSTFPLIDSSSQNQIRKRIVLEKLSK VLPGLLVPLQITLGDITYT
2463	C	135	341	MYTIKIKPRSFGIIHNLPSKPGPLFLPHSLIGWFDFTA SFLYPMNCSAMHHXVRKSSSATAITKIGKTG
2464	A	265	395	RLCDGLFPQQDPAAPAPCEETQLSLLPLQGCGLMEGK TMEAKT
2465	A	88	1496	QETSKMETLSFPRYNVAEIVIHNRNKILTADGKNLT KNDLYPNPKPEVLHMIYMRALQIVYGIRLEHFYMPV NSEVMYPHLMEGFLPFSNLVTHLDSFLPICRVNDFET ADILCPKAKRTSRFLSGIINFIFREACRETYMEFLW QYKSSADKMQQLNAAHQEALMKLERLDSVPVEEQEEF KQLSDGIQELQQSLNQDFHQKTIVLQEGNSQKKSNI EKTKRLNELKLSVSVSLKEIQESLTKTIVDSPEKLKNY KEKMKDTVQKLKNARQEVVEKYEIYGDSVDCLPSCQL EVQLYQKKIQDLSDNREKLASILKESLNLEDQIESDE SELKKLKTEENSFKRLMIVKKEKLATAQFKINKKHED VKQYKRTVIEDCNKVQEKRGAVYERVTTINHEIQKIR LGIQQLKDAADREKLKSQEIFLNLKTALEKYHDGIEK AAEDSYAKIDEKTAELKRKMFKMST
2466	A	194	2287	GMGSENSALKSYTLREPPFTLPSGLAVYPAVLQDGKF ASVFVYKRENEDKVNKAAKVP**HLKTLRHPCLLRFL SCTVEADGIHLVTERVQPLEVALETLSAEVCAGIYD ILLALIFLHDRGHLTHNNVCLSSVFVSEDGHWKLGGM ETVCKVSQATPEFLRSIQSIRDPASIPPEEMSPEFTT LPECHGHARDAFSFGTLVESLLTILNEQVSADVLSSF QQTLHSTLLNPIPKWRPALCTLLSHDFFRNDFLEVVN FLKSLTLKSEEEKTEFFKFLLDREVSLSEELIASRLV PLLLNQLVFAEPVAV\KSFLPYLLGPKKDHAAQGETPC LLSPALFQSRVIPVLLQLFEVHEEHVRMVLLSHIEAY VGALSLREQLKKV\IL\PQVLLG\LRD\TSDSIVAIT LHSLAVLVSLLGPEVVGGERTKIFKRTAP\SFTK\N TDLSLEGDPFSQPIKFPINGLSDVKNTSEDSNFSS SKKSEEWPDWSGPE\EPENQTVNI\QIWP\REP\CDD VKSQCTTLDVEESSWDDCEPSSLDTKVNPGGGITATK PVTSGEQKPIPALLSLTEESMPWKSSLPQKISLVQRG DDADQIEPPKVSSQERPLKVPSELGLGEEFTIQVKKK PVKDPMDWFADMIPEIKPSAAFLILPELRTEMVPPK DDVSPVMQFSSKFAAAEITEGEAEGWEEEGELNWEDN NW
2467	A	2	868	IAGVAVFFYRDMFVRKDRKIHKDAESAQSCTDSSGSF AKLNLGLFDSPVKEYQQNIDSPKLIVT/SLTSRKELPP NGDTKSMVMDHRGQPPELAALPTPESTPVLHQKTLQA MKSHSEKAHGHGASRKETPQFFPSSPPPHSPLSHGHI PSAIVLPNATHDYNTSFSNSNAHKAEEKLQNIHDHPLT KSSSKRDHRRSVDSRNTLNDLLKHLNDPNSNPKAIMG DIQMAHQNLMLDPMGSMSEVPKVPNREASLYSPST

681
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LPRNSPTKRVDVPTTPGVPMTSLERQRGYHK
2468	A	483	764	MGSVFWHVLFCISGVCLWCAHRMAAFLQQMAVLLPVD CERPAAVHWLALCGCCYQQLVWESRTRSCFWSLECLC FGGQHFGSVPSFFCSSVWL*
2469	A	3	357	FGFNGCSKRRIKLOELSDLEERENEDSMVPLPKQSLK FFCALEVVLPSCDCRSPGIGLVVEEPMDKVEEGPLSFL MKRKTAQKLAIQKALSDAFQKLLIVVLG/QDCLDHP* STSVSVSK
2470	A	3	57	RIGQGVVPHS*VEGGPNVISIVLEYLRDTPVPVHV CDGSGRASDILAFGHKYSEEG*VKVFLWCTHKWKED PM
2471	A	69	512	MALAFGLTVLSKATLGARLTTHCAHPARRARAFSSDV MTHSSILTRASLLTLWTMFTTRTKILTEGSGVSWWAA AFPRDVVAGGSILALASLMTVVTIGALLTAVLAAPAP EARSTVASPGDGVAQSPIFALAPAGAVGTPVITIAG*
2472	A	2195	872	VSQATDVEVGTDLVPSVTVKVTLQNRVILQAKLSVY VQPPLELTCDQFTFEFMNRNPDGIPRVIQCKFRLPLK LILCLPGQPSKTASHKITIDTNKSPVSLLSLFPGFASQ SDDQVNVMGFHLG\GAR\ITVLASKTSSTDIRIPG VEQFE\DLWASLTNELILRLQEYFEKQGVKDFACFS G\SITPFKEYF\ELIGSIHFELRINGEKLEELLSERA VQFRAIQRRLLARFKDKTPAPLQHLGHLVRMGTYK\Q VIALA\DAVGKTKGNLFQSFTRLKSATHLVILLIAL WQKLSADQVAILEAAFLPLQEDTQELGWEETVDAEIF H\L*KTCCRKSAKQQALNPPGRLTYPNDTS\QLKKHI TLLCDRLSKGGRCLCLSTDA/APHQTMVMPGGCTTIP ESDLEERSVEQDSTELFTNHRHLTAETPRPEVSPLOQ VSE
2473	A	1	473	EVRWNSPPTDSLSPDGGSIELEFYLAPEPFMSPLL APPYSGLGVDGPYAPLMVLMCRVCLEDKPIKPLPCC KKAVCEECLKVYLSAQIQCPCTCQFVWCFKCHSPWHEG VNCKEYKKGDKLLRHWASEIEHQRNAQKCPCKKIHI QRTGCDHM
2474	A	131	1098	RVPAGGARRLGQDPPRLPPGVADAPAAMSTQRLRNED YHDYSSTDVSPEESPSEGLNNLSSPGSYQRFQSNST TWFQTLIHLLKGNIGTGLLGLPLAVKNAGIVMGPI LIIGIVAVHCMGILVKCAHHFCRRLNKSFVDYGD YGLLESSPCSWLRNHAHWGRRVVDFFLIVTQLGFCCVY FVFLADNFKQVIEAANGTTNNCHNNETVILTPTMDSR LYMLSFLPFLVLLVFIRNLRALSIFSLLANITMLVSL VMIYQFIVQIL*MDLQPM*QTKVFHREQVPLCLQHVE SQMEQFWAECFAQRVLPINVLSLQKK
2475	A	131	1098	RVPAGGARRLGQDPPRLPPGVADAPAAMSTQRLRNED YHDYSSTDVSPEESPSEGLNNLSSPGSYQRFQSNST TWFQTLIHLLKGNIGTGLLGLPLAVKNAGIVMGPI LIIGIVAVHCMGILVKCAHHFCRRLNKSFVDYGD YGLLESSPCSWLRNHAHWGRRVVDFFLIVTQLGFCCVY FVFLADNFKQVIEAANGTTNNCHNNETVILTPTMDSR LYMLSFLPFLVLLVFIRNLRALSIFSLLANITMLVSL VMIYQFIVQIL*MDLQPM*QTKVFHREQVPLCLQHVE

682
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				SQMEQFWAECFAQRVLPINVLSLQKK
2476	A	505	1373	WGDGTQNESHSVSLTAFLSDTKDRGPPVQSQIWSR GEKVPFVQTYSLRAFEKPPQVQTQALRDFEKHLNDLK KENFSLKLLIYFLEERMQQKYEASREDIYKRNTTELKV EVESLKRELQDKKQHLDTWADVENLNSQNEAELRRQ FEERQQEMEHVYELLENKMQLLQEE SRLAKNEAARMA ALVEAEKECNLELSEKLKGVTKNWEDVPGDQVKPDQY TEALAQRDK*VPSVLFL\RLSFAHSQGIQQLSCSLSR T/RQ*ELHYF*DFMGPPQKTFFSGLNFQWYPL
2477	A	1	317	QRPSEAKEIKLYAQIPPIEKMDASLSMLANCEKLSLS TNCIEKIANLNL\EAVGDTLEELWISYNFIEKLGKI HIMKKLKILYMSNNLVKDWGTPVIKGDEEEDN
2478	A	2	607	CKNTLIRQNI PRAQFPATSPRSIIQQPN/PFPRRFVL PLNVSLNAPEGDNLSPLSYTSASAVKQADGTIWC SHE NLHQEDLEKEGGIEFPQIYYDRFSGKKYHFFYGCGFR HLVGDSLKVDVVKNTLKVWREDGFYPSEPVFVPAPG TNEEDGGVILSVVITPNQNESNLLVLDKKNFEELGR AEVPVQMPYGFHGTFIPI
2479	A	2	607	CKNTLIRQNI PRAQFPATSPRSIIQQPN/PFPRRFVL PLNVSLNAPEGDNLSPLSYTSASAVKQADGTIWC SHE NLHQEDLEKEGGIEFPQIYYDRFSGKKYHFFYGCGFR HLVGDSLKVDVVKNTLKVWREDGFYPSEPVFVPAPG TNEEDGGVILSVVITPNQNESNLLVLDKKNFEELGR AEVPVQMPYGFHGTFIPI
2480	A	101	580	LSLTKNCALLGEETMMEQEMTRLHRRVSEVEAVLSQK EVELKASETQRSPLQDLATYITECSSLKRSLEQARM EVSQEDDKALQLLHDIREQSRKLQEIKEQEYQAQVEE MRLMMNQLEEDLVSARRRSDLYESELRESRLAAEEFK RKATECQHKLLK
2481	A	1	2025	MAWAGRGRGSRQGSSELHLPWAIDVCLFSLVRSRGRFL REVWWEIWKVLLLLHVANGAQQAGPIPWNTGLQANH SVPVSKPHQKWPVQHFQELLRSANSLTAPFKQVQYWR GTKMNQRPVPQIHSWFRMFCGMAHESHGIGKWGVAL EGHPPGPGKQESIANACWEAAVRSPGSRSHKAETKSS KSRDQILSVLRPASFVRDKSI PQPWLESDGINKRWSP TCLSGEPSLGRVNPLLHELQTQCFVRTPSYQRATEAA KPQERCTIQLNKMCCLOAGSFSRYASVIAIKHICHAH STPKALLTSFLVLTTRSLNLHLHLRLSHDPDKFRDGG VSSSQYSRYCSLTQPDFDSSNSSTFFLLLTISLLSSQ FCIRLISLPECPVSQWQEAAREHLGGGSDLSSMGETH PDLGGGPSEGPGGWPEQVSAFAQLVLVSTMSFQGT WRKRFSSTDQILPFTCAYGLVLQVPMMHQTTEVNYG QFQDTAGHQVGVLELPYLGSVSLFLVLPDRDKDTPLS HIEPHLTASTIHLWTTSLRRARMDVFLPSELTKPEFR WDQRLFALVLRPLPGTMSVESEQLTGVPLDDSAITPMC EVTGVGMECFSDAKDTIEDLSEMHSQDLSEMRGNPT KPSPPLSGTTVENFGSRGTDSEAFSEPSLGKEPVTH RTRVPLQWP
2482	A	137	879	LPPRGPATFGSPGCPPANSPPSAPATPEPARAPERVM ANAGLQLLGFILAFLGWIGAI VSTALPQWRIYSYAGD

683
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				NIVTAQAMYEGGLWMSCVSQSTGQIQCKVFDSSLNLSS TLQATLALMVVGILLGVIAIFVATVGMKCMKCLEDDE VQKMRMAVIGGAIFLLAGLAILVATAWYGNRIVQEFY DPMTFVNARYEFGQALFTGWAAASLCLLGGALLCCSC PRKTTSYPTPRPYPK\ PAPS\ SGKDYV
2483	A	200	1139	RIISTITYQFSAALGQEVFYITFLPFTHWNIDPYLSR RLIIIIWVLVMIYIGQVAKDVLKWPRPSSPPVVKLEKRL IAEYGMPSHAMAATAIAFTLLISTMDRYQYPFVLGL VMAVVFSTLVCLSRLYTGMHTVLDVLGGVLITALLIV LTPAWTFIDCLDSASPLFPVCVIVVPFFLCYNYPVS DYSPTRADTTTILAAGAGVTIGFWINHFFQLVSKPA ESLPVIQNI PPLTTYMLVLGLTKFAVGIVLILLVRQL VQNLSLQVLYSWFKVVTNRNKEARRRLEIEVPYKFVTY TSVGIGTKVVAQMPTDV
2484	A	173	307	SHICLKKSASLTGTWMKLETIILSKLTQEQKTKHCM FSLISGS
2485	A	173	307	SHICLKKSASLTGTWMKLETIILSKLTQEQKTKHCM FSLISGS
2486	B	86	225	PRQEKSSHVSTRRSPKLLREKPEAAAGEAAAEAGLP MFARSRRSR
2487	A	14	1256	WPCGAAPGLTHASERMFTLTMTMIQALAPVMGWDRKPL KMFSSEEMRGHLHHHKKCLTKILKVEGQVPDLPSCLP LTDNTRMLASILINMLYDDLRCDFPERDHFRIKICEEYI TGKFDPPQMDKNLNAIQTVSGILQGPFDLGNQLLGLK GVMEMMVALCGSERETDQLVAVEALIHASTKLSRATF IITNGVSLKQIYKTTKNEKIKIRTLVGLCKLGSAGG TDYGLRQFAEGSTEKLAKQCRKWLCNMSIDTRTRWA VEGLAYLTLDADVDDFVQDVPALQAMFELAKTSDKT ILYSVATTLVNCTNSYDVKEVIPELVQLAKFSKQHPV EEHPKDKKDFIDMRVKRLLKAGVISALACMVKADSAI LTDQTKELLARVFLALCDNPKDRGTIVAQGGGKALIP LALEGTD
2488	B	526	3482	MDSLKQETQGLQKEKESREKELMGFSKSVNEARSKMD VAQSELDIYLSRHNTAVSQLTKAKEALIAASETLKER KAAIRDIEGKLPQTEQELKEKEKELQKLTQEETNFKS LDKMAVWAKKMTETIQTTPENTPRLFDLVKVKDEKIRQA FYFALRDTLVADNLDQATRVAYQKDRRWVVTLQGOI IEQSGTMTGGGSKVMKGRMGSSLVIEISEEEVNKMES QLQNDSSKAMQIQEQKVQLEERVVKLRHSEREMRNTL EKFTASIQRLEIEQEEYLVNVQVKELEANVLATAPDKKK QKLLEENVSAFKTEYDAVAEKAEEESLPEIQKEHRNLL QELKVIQENEHALQKDALSILKLEQIDGHIAEHNSK IKYWHKEISKISLHPIEDNPIEISVLSPEDELAIKN PDSITNQIALLEARCHMKNLGAIAEYKKKEELYLQ RVAELDKITYERDSFRQAYEDLRKQRLNEFMGSRPP KKSWKKIFNLSGGEKTLSSLALVFALHHYKPTPLYFM DEIDAALDFKNVSIVAFYIYEAVWFLSNITAGNQOQV QAVIDANLVPMIIHLDDKGDFTQKEAAWAISNLTIS GRKDQVAYLIQQNVIPFCNLLTVKDAQVQVQVLDGL SNILKMAEDEAETIGNLIEECGGLEKIEQLQNHENED

684
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				IYKLAYEIIDQFFSSDDVFNNFVFLILYINIDLKVIYIGSSLMKQTPSPGVTVYESVVVVLVYNACKLDSPSGG
2489	A	1	747	MRLQRPRQAPAGGRRAPRGGRGSPYRPDPGRGARRLR RFQKGEGAPRADPPWAPLGTMALLALLVVALPRVW TDANLTARQRPEDSQRTEGDNRVWCHVCERENTFE CQNPRRCKWTEPYCVIAAVKIFPRFFMVAKQCSAGCA AMERPKPEEKRFLLLEPMPFFYLKCKKIRYCNL/GGA /NLSTHQ\CSKNMLGAWVRAVVGCGWPSSCCWPPLQP ASACLEPRDCHRLSLPEHGLAPDRCHLLH
2490	A	2	1177	GFVEAGEECYCVS\GQECRDLCCFAHNCSLRPGAQCA HGDCCVRCLLKPAGALCRQAMGDCDLPEFCTGTSSHCPD VYLLDGSPCARGSGYCWGACPTLEQQCQQLWGP GSHPAPEACFQVVSAGDAHGNCGQDSEGHFLPCAGR DALCGKLQCQGGKPSLLAPHMVPVDSTVHLDGQEVTC RGALALPSAQLDLLGLGLVEPGTQCGPRMVCQSRRCR KNAFQELQRCLTACHSHGVCNSNHNCHCAPGWAPPFC DKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLPLLPG AGLAWCCYRLPGAHLQRCSWGCRDPACSGPKDGP HR DHPLGGVHPMELGPTATGQPWPLDPENSHEPSSHPEK PLPAVSPDPQADQVQMPRSCLW
2491	A	1	609	AAARTFWYKLFPCRGSAGAAKAAEQKRQVGGRAEPGT AAPCGARCPGPTPGWQVPATKALLSQPMGCPPPGPCR GHT*ADPQLPLTHAP/PEARLSPPQPP/PSPPGSATP GA*AGVASPKPTLPAPGAPGTPQRLPGP/RREKPAFL SQPESST*PHPTPVSAASSSPA/PESSCHDELGLLSL NLPAPGPPKPTPGAAASFQSG
2492	A	1	242	MNRGGFAVKILALLDALSTVCSQRVQKAKKQOHLQNK EHFKALLKQKEKLKQOEDL/RKKLF*IQGIRCPQATP HHGQCSL
2493	A	909	353	RSFVLDTASAI CNYN AHYKNHPKYWCRGYFRDYCNII AFSPNSTNHVALRDTGNQLIVTMSCLTKEDTGWYWCG IQRDFARDDMDFTELIVTDDKGTLANDFWSGKDL SGN KTRSCAPKVVVRKADRRTSILIIICILITGLGIISVI SHLTKRRTSRQNRNRVGNLTKPF SRVLT PKEMAPTEQM
2494	A	516	848	MWSLWIWVDQHARLIPSPQVLLLLLLRETPSTAAAVA GWLVVASMALLQLHAVGGVALTSSHPFMWATGEELRK PPWQGSAGSASGVEELTGKHSCPGPEEPATVQKAPA*
2495	A	349	1018	TFTQPD PDDLISKPPRTPGGG*YQTQWSPDPDRRTS PAGRPGPARRPPRRTPRPARGRHPGR*GGPGASRPGG TGAAPAADQTGSPAVSTPSEFGAPGQAEGPQSPIRAS ARSHLSCTAWLGKPSKPSAQRQPTVGPDRDGDSSQA PNLSRGQAWRASLASPQNTSATGRVTCHGQSTWPLCR LKSNRRRKSGFA/GNKSEPVGLTRRSKHQPRNPQGQV GI
2496	A	349	1018	TFTQPD PDDLISKPPRTPGGG*YQTQWSPDPDRRTS PAGRPGPARRPPRRTPRPARGRHPGR*GGPGASRPGG TGAAPAADQTGSPAVSTPSEFGAPGQAEGPQSPIRAS ARSHLSCTAWLGKPSKPSAQRQPTVGPDRDGDSSQA PNLSRGQAWRASLASPQNTSATGRVTCHGQSTWPLCR

685
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				LKSNRRRKSGFA/GNKSEPVGLTRRSKHQPRNPQGQVGI
2497	A	349	1018	TFTQPDPPDDLISKPPRTPGGG*YQTQWPSPPDPRTS PAGRPGPARPPRRTPRPARGRHPGR*GGPGASRPGG TGAAPAADQTGSPAVSTPSEFGAPGQAEQPQSPIRAS ARSHLSCTAWLGKPSKPSAQRQPTVGPDPGDRDGSSQA PNLSRGQAWRASLASPQNTSATGRVTCHGQSTWPLCR LKSNRRRKSGFA/GNKSEPVGLTRRSKHQPRNPQGQVGI
2498	A	2025	422	PPGTQGSPPQT/GDHGGKPPLPAEKPAPGPGPLPARAS RAEGRGASGWKPGGQPAGGSWQGGDAGPRRPASGDQR TAGAAKALAGPAGEAAGGDRGAAQGDPPAEAGGRGG* TQAGGGASRARGSGAQRPGGP*RQGQDGGESASPAF GPCPQSSWGPPCSI PGP*PALPGAL*GA\VG RDPAGP PDGGPDTEP/PGSPGQAERWPEGCRPQGSWHCEGAPQ GPGAGARARPRQGSRGPRGAPRRGI PWAKSGR\TGGS QDRKKPGKEVAATGTST/PEGSQ LARGRARSRDGGPS HEAQASEPRPGPCSGPARWGGRSSCTAPGCVTPAGTA GHL*WRAGWTAGPPAGPWRS PGDEKGP RGGPCACVPR AAERRGGRCPCGAQAEARARAGAQTSCPGGPEAGQCQ AQP GPETAGWLRPPEATAGPWPS CRGSAGPEGWGHWW P*PPA*CPGERPPWRPGCPAPPGCCGSSAGGPQPAA* TGAWASRGVLAPAGHEGHASHCPRPAAGLSQPHPSQ TLEVTLAS PQGFMSEALTKCE
2499	A	1415	661	SLRTPGFRGGGVLYWDAGAAGTGSNHALGANVELWIM LLQVVREGKFSGFLTSCSLLLPRAAQILAAEAGLPSS RSFMGFAAPFTNKRKAYSERRIMGYSMQEMYEVVSNV QEYREFVPWGKSLVSSRKHLKAQLEVGFPPVMER YTSAVSMVKPHMVKAVCTDGKLFNHLETIWRFSPI AYPRTCTVDFSI SFEFRSL LHSQ LATMFFDEVVKQNV AAFERRAATKFGPETAI PRELMFHEVHQT
2500	A	673	941	CCLAAHSGPPAQGQRRGPG*LCCSAGSGGNL*S*AGG PG*GRSGQPVCPPWP GPGAPGHRPALPGSGGSSAVGR SAVPGAVRSPSHAGW
2501	A	328	1212	RQEQGHFHFFCGGMSSFKAGTSHLDVYMQVTEGREDY NPSMHLAKRQFLSLEEEAEDYNPSQHRAQGNWLQDYN ASMQRVHGQCVSLEEDVELCVPRWACREMQSHNYP LVAGLQQYNFSISLAQGECTSHWRKRGIMTYSSIHCL GDVTLHSYLGPSKTEDCDISVTLPRLERRITLPKHW IKKYFTIFLMGKAQINKIDRPLVRQIKEKREKNQORDA IKNDKGDITTKPTEIQTTIREYYKHLYGNKVENLEEI DKFLETSTPPRLNEEEVESLSRPIAGSEIEALINSL
2502	B	1	1428	MGSRVRLSKRRKAGVQSGTNALLVVKHRDMNEKELE AQEARKAQLNHEPEEEEEEEIRQPRKKLGAQPVVHW VAPDGRLLGNSSRTRVRGDGTLDTITTLRDSGTFTC IASNAAGEATAPVEPRGLCPDYACTRFSTTVPLMTPS STGVDIEAARKEEERIMLRDARQWLNSGHINDVRHAK SGGTALHVAAAKGYTEVLKII SLRFGVPRTQVRTWVA LYEKHGEKGLIPKPKGVSADPELRIKVVKAVIEQHMS LNQAAAHFMLAGSGSVARWLKVYEERGEAGLRALKIG

686
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				TKRNIAISVDPEKAASALELSKDRRIEDLERQVRFL TRLMYLKKLKALAHPTKKAAEI PRSTFYHLKALSKP DKYADVKKRI SEIYHENRGRYGYRRVTL SLHREGKQI NHKAVQRLMGTLSHKAAIKVKRYRSYRGEMKKLRIRE VQILAGGHTAKLNMEQVKSADAFTYIKQPIA
2503	A	218	415	MRCRAPAWLRRLCGQLLSERLMRPNGVQAVVRGILEG AGAGAAGGSDAEVTAAADWKKCDLIAKILA
2504	A	3	136	SWATAGAANGPAPLGVRAPPARTSPAEMGATGAEE PLQSVLWVKQQRCAVSLEPARALLRWWRSPGPGAGAP GADACSVPVSEIIAVEETDVHGKHQSGSKWQKMEKPY AFTVHCVKRARRHRWKWAQVTFWCPEEQ LCHLWLQTL REMLEKLTSRPHLLVF INPFGGKGQ GKRIYERKVAP LFTLASITTDIIVTEHANQAKETLYEINIDKYDG*VR RPSASARPQPGGRARRRRWGRGRSRRCNPCCG
2505	A	335	1105	MKRERGA LSRA SRA LRLAPFVYLLLIQTDPLEGVNIT SPVRLIHGT V GKSALLSVQYSSTSSDRPVVKWLKRD KPVTVVQSIGTEVIGTLRPDYRDRI RL FENGSLLLSD LQLADEGTYEVEISITDDTFTGEKTINLTVDPISR QVLGASTTVLELSEAF TLNCSHENGTKPSYTWLKD GK PLLND SRMLLSPDQKVLTI TRVLMEDDDLYSCVVENP INQGR TLPC KITEYRKSSLSSIWLQEA FSSLGPW*
2506	A	335	1105	MKRERGA LSRA SRA LRLAPFVYLLLIQTDPLEGVNIT SPVRLIHGT V GKSALLSVQYSSTSSDRPVVKWLKRD KPVTVVQSIGTEVIGTLRPDYRDRI RL FENGSLLLSD LQLADEGTYEVEISITDDTFTGEKTINLTVDPISR QVLGASTTVLELSEAF TLNCSHENGTKPSYTWLKD GK PLLND SRMLLSPDQKVLTI TRVLMEDDDLYSCVVENP INQGR TLPC KITEYRKSSLSSIWLQEA FSSLGPW*
2507	A	1160	3149	VSKTTTTNAGNALFPM PGSSKTKKPN SHQRGQMGS*G RNPPSLGRAPAPLPEREAPI PAPQLGPSAAGTSRQVG QKSSTSPHQGE EAILNRELKKKDGKKK*KK/PTGLSK HQPAGFIQNE*NLKGAGEFVQGLAGSQNPSSKLQGL GG\SAESRGFSRGQQTAPHWESTPLKGALPPC PERG MLPEEG*GFSGKEASSGPVQPQPTCLYGIRPSLGS*P *GQRRTLLAPTFLQENQL\SGPSPGQRARSVLRPFSA /PGLRPELELTGGRGSTRSRAAGPWASDCTAGSDQE SLGRSSGKGR*GASGTVLG VSMCKV/PGCKAAGGHL P GGGRGLDLECGWGLRSWLPGRGRQ/TGPPG/PQGRDS *STKQSDSHRWQDSGGGLAPPPPGQGNNGARPCC*DV TKASAPGVSGDTGREAPSATGISTFRSCCMSSARGLG QSPAAPVLASSFLPTSCTGPPGLPGLPSSGSEENIHS GAWALVGQEGPSMDGRGNMMLRGVWTGVHGGGMD\G CGAEVI*RGKFLME*YRSLQRKQDSSPARTPAPQWL SITTGS*TPE/GDPGGKLDAAQRGRAIAAH/GTAGGC CPRCCCHL*SPGSARSSP/PMASASIRVS\PPRSGGS PPSPSSA*KSDRTDAGAGVAAAASPGAGAPAHCPQGP PRSCQGPQRR
2508	A	1	957	METSSPRPPRPSSNPGLSLDARLGVDTHLWAKVLF LYALI WALGAAGNALSVHVVLKARAGRAGRLRHVLS LALAGLLLLLVGVPVELYSFVWFHYPWVFRDLGCRGY

687
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				YFVHELCAAYATVLSVAGLSAERCLAVCQPLRARSLLT PRRTRWLVALSWAASLGLALPMAVIMGQKHELETADG EPEPASRVCTVLVSRTALQVFIQEAIIVMYVICWLPY HARRLMYCYVPDDAWTDPLYNFYHYFYMTNTLFYVS SAVTPLLYNAVSSSRKLFLEAVSSSLCGEHHMPKRLP PKPQSPTLMDTASGFGDPPETRT
2509	A	144	291	DVEVKWIEYQNMVNYLIQWIRHHVTTMSERTFPNNPV ELKVSVTVEIT
2510	A	144	291	DVEVKWIEYQNMVNYLIQWIRHHVTTMSERTFPNNPV ELKVSVTVEIT
2511	A	3	279	RSLPPAHSVSLWSVKDGLRPWHPELRSVQPTRGGRTQ THRRGAAPGISTPHTLGGRASAARRPWHTCGRQRRPP RRRRERRPLYSSVLRST
2512	A	3	1396	RQENNTRGVPSLLKSFLQERLGIHLIRRKIVKPKHHV LMSRKESWKVKSEIPKVPKQPLVLHHPMTTTSKSPSK DMLEPEAELEAEDLPTTKSTSVES/EDAH*EPGRPFV LPDL/PCHCLPSAPTPLCIVKRPCPT*VTQLSASAQS AHQMRTTPRAQSPSS*PR*VNCLPPS/LHKDDLELKEK DQKKPPTAPREVKGTRRKLPTAFLPSKYHGYEELLTA KPDPAFIEPKGIQKNA/PSPATNAEAPTVPPLLQAQA GHSSETLCSQRETGPENPDSTPKED*SPTSG*HLHSL AGSPEHYRGSTRCCPAPVDRTAAGEP/ASSTWRPRGC *RSSRHVTGSW*VALCAQCSGLPRSPWPAQR*VRASP SSATSSSSWMSSARSPQPVTHKARAVHGGCVHHPACA PALPEGSVPWTAPQG*PAGHRPQSSAGPHLLATRWHP LVRISPPWPRHDLVPGPAAIKSGCTGQ
2513	A	3	1396	RQENNTRGVPSLLKSFLQERLGIHLIRRKIVKPKHHV LMSRKESWKVKSEIPKVPKQPLVLHHPMTTTSKSPSK DMLEPEAELEAEDLPTTKSTSVES/EDAH*EPGRPFV LPDL/PCHCLPSAPTPLCIVKRPCPT*VTQLSASAQS AHQMRTTPRAQSPSS*PR*VNCLPPS/LHKDDLELKEK DQKKPPTAPREVKGTRRKLPTAFLPSKYHGYEELLTA KPDPAFIEPKGIQKNA/PSPATNAEAPTVPPLLQAQA GHSSETLCSQRETGPENPDSTPKED*SPTSG*HLHSL AGSPEHYRGSTRCCPAPVDRTAAGEP/ASSTWRPRGC *RSSRHVTGSW*VALCAQCSGLPRSPWPAQR*VRASP SSATSSSSWMSSARSPQPVTHKARAVHGGCVHHPACA PALPEGSVPWTAPQG*PAGHRPQSSAGPHLLATRWHP LVRISPPWPRHDLVPGPAAIKSGCTGQ
2514	A	1065	478	HGLCELTSTVQEGELCVFFRNHFFSTMTKYKGQLYLL VTDQGFLTEEKVWVESLHNVDGDGNFCDSEFHLRPPS DPETVYKGQDQIDQDYLMAISLQQEQQSQEINWEQI PEGISDLELAKKLQEEEDRRASQYYQEQAIAAAAAA ASTQAQQGQPAQASPSGRQSGNSERKRKEPREKDKE KEKEKNSCVIL
2515	A	1065	478	HGLCELTSTVQEGELCVFFRNHFFSTMTKYKGQLYLL VTDQGFLTEEKVWVESLHNVDGDGNFCDSEFHLRPPS DPETVYKGQDQIDQDYLMAISLQQEQQSQEINWEQI PEGISDLELAKKLQEEEDRRASQYYQEQAIAAAAAA ASTQAQQGQPAQASPSGRQSGNSERKRKEPREKDKE

688
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				KEKEKNSCVIL
2516	A	290	1041	KACLHLLSSFLTSTNLFNPLLPDSLVSVEARSQRANL GPCRRKRLQTLMLRLAAGFQYSSHKDPSLSAKEKHTDY HNEARGPWPGWVG*RTADGSCGRGPDGAHHPGPKSSS WRASRLLPGLGGSHHLDAYVGRDLECGTPAPLQLEIP PQPRGHPAPIPTGQAGPRDSGPGASP*VETRPLTDGR R*PGVRPVGWTPAHPAGTLRPRGAVEPSVSACGKWAP SPTSQGCCEGRCDVPAKHPRAWRTPLCSQ
2517	A	2	1736	QNENSVDKWGKPLVIDKLEMAKVEGLWNLFLPAVSG LSHVDYALIAEETGKCFAPDVFNCQAPDTGNMEVLH LYGSEEQKKQWLEPLLQGNITSCFCMTEPDVASSDAT NIECSIQRDEDSYVINGKKWWSSGAGNPKCKIAIVLG RTQNTSLSR*LNNSD*ETCVGMSQSSSYLGNLLKIH LDSQIIM*DMRVNVIYLYFTSIF*QVFLENIIGSIAE HSSLWNFQY*KVLLNYQSCLD*IIRQIFSDLCNEVIR CLDQRQ*S*NV*LYI*VPSYHC*AVRSFNQTTHLFSN HCFCSRSQPASDYVGVRLHSSSHSHHCLHDYMKTSK RQLGFCLLSVLFFFLANFF*YNFSFD*\HKQHSMILV PMNTPGVKIIIRPLSVFGYTDNFHGGHFEIHFNQVRVP ATNLILGEGRGFEISQGRLGPGRIHHC MRTVGLAERA LQIMCERATQRIAFKKKLYAHEVVAHWIAESRIAIEK IRLLTLKAAHSMDTLGSAGAKKEIAMIKVAAPRAVSK IVDWAIQVCGGAGVSQDYPLANMYAITRVLRLADGPD EVHLSAIAATMELRDQAKRLTAKI
2518	A	2	1736	QNENSVDKWGKPLVIDKLEMAKVEGLWNLFLPAVSG LSHVDYALIAEETGKCFAPDVFNCQAPDTGNMEVLH LYGSEEQKKQWLEPLLQGNITSCFCMTEPDVASSDAT NIECSIQRDEDSYVINGKKWWSSGAGNPKCKIAIVLG RTQNTSLSR*LNNSD*ETCVGMSQSSSYLGNLLKIH LDSQIIM*DMRVNVIYLYFTSIF*QVFLENIIGSIAE HSSLWNFQY*KVLLNYQSCLD*IIRQIFSDLCNEVIR CLDQRQ*S*NV*LYI*VPSYHC*AVRSFNQTTHLFSN HCFCSRSQPASDYVGVRLHSSSHSHHCLHDYMKTSK RQLGFCLLSVLFFFLANFF*YNFSFD*\HKQHSMILV PMNTPGVKIIIRPLSVFGYTDNFHGGHFEIHFNQVRVP ATNLILGEGRGFEISQGRLGPGRIHHC MRTVGLAERA LQIMCERATQRIAFKKKLYAHEVVAHWIAESRIAIEK IRLLTLKAAHSMDTLGSAGAKKEIAMIKVAAPRAVSK IVDWAIQVCGGAGVSQDYPLANMYAITRVLRLADGPD EVHLSAIAATMELRDQAKRLTAKI
2519	A	2	550	FGVINLICTGFLLMWCSSTNSIALT\SYTYLTIFDLF SLMTCLISYWVTLRKPSPVYSFGFERLEVLAFASTV LAQLGALFILKESAERFLEQPEIHTGRLLVGTFFVALC FNLF TMLSIRNKPFAYVSEAASTSWLQEHVADLSRSL CGIIPGLSSI FLPRMNPVFLIDL A GAFALCITYML
2520	A	1	1876	RAPMMTKAVPEEPRKPGRLTQALNSPLTWEHVWICVP GGTPDCLTD TFRVKRPHLRRSASNGHVP GTPVYREKE DMYDEIIELKKS LHVQKSDVDLMRTKLRRLEEENSRK DRQIEQLLDPSRGTD FVRTLA EKRPDASWVINGLKQR ILKLEQQCKEKDGTISKLQ TDMKTTNLEEMRIAMETY

689
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				YEEVHRLQTLASSETTGKKPLGEKKTGAKRQKKMGS ALLSLSRVQELTEENQSLKEDLDRVLSTSPITISKTO GYVEWSKPRLLRRIVELEKKLSVMESSKSHAAEPVRS HPPACCLASSSALHRQPRGDRNKDHERLRGAVRDLKEE RTALQEQLLQORDLEVQQLLQAKADLEKELECEGEE ERREREEVLREEIQTLTSLQELQEMKKEEKEDCPEV PHKAQELPAPTSSRHCEQDWPPDSSEGLPRPRSPC SDGRRDAAARVLQAQWKVYKHKKKKAVLDEAAVVLQA AFRGHLTRTKLLASKAHGSEPPSPVGLPDQSSPVPRV PSPIAQTGSPVQEEAIVIIQSALRAHLARARHSATG KRTTTAASTRRRSASATHGDASSPFLAALPDPSPSG PQAVAPLPGDDVNSDDSDDIVIAPSLPTKNFPV
2521	A	5618	4060	APARRGLGDRCSSSSSFSSSFSSASSPRRLATAAARA GGAAVIPVPEEPALPVPGGRGAGEAGPRRTQQVEPGV PGRAPPAHHAALCHLSRPQAKILSMMEDNKQLALRID GAVQSASQEVTNLRAELTATNRRRLAELSGGGGPGPGP GAAASASAAGDSAATNMENPQLGAQVLLREEVSRLQE EVHLLRQMKEMLAKEESQGGKSSEVLSETELRLVQL AQKEQELARAKEALQAMKADRKRLKGEKTDLVSQMQQ LYATLESREEQLRDFIRNYEQHRKESEDAVKALAKEK DLLEREKWEELRRQAKEATDHATALRSQDLKDNRMKE LEAELAMAKQSLATLT KDVPKRHSLAMPGETVLNGNQ EWWVQADLPLTAAIRQSQQTLYHSHPPHPADRQAVRV SPCHSRQPSVISDASAAEGDRSSTPSDINSRHRTHS LCNVRPAAAGPGPLGPAQKLQGRGWRGEAILAVSSRP PREHSGECISCSVLSFCKKRWMWGEKGMRPVCSLCPG G
2522	A	1023	766	MLCSRLGTTASWRLGIRAWAPLLLLFPWDWHFILSF SSRPWAGTLLAPHDVIMGSSTFPQSCQAEAGPRHAWP TGRFSRRLRRV*
2523	A	1	429	NTLLTIIVLFPDPPSLSSNSSIRSSSSSFSTCISCELS TSGCPAITTESVSASPSMISPSATSV*VTS*SSSCTS ASPGSPGSCWLLLES*EAPWASCSDLFLLLEALLLPKR LLGWFTIRESVSKGFRAALTVLAMLGLDRSKL
2524	A	165	638	MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLA LALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHS DLM CTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGK LHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYK HHWQGPFW*
2525	A	165	638	MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLA LALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHS DLM CTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGK LHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYK HHWQGPFW*
2526	A	2	266	KGSTEAFIGTAGWGTGLLPSSAGLPGGWGPAGGWAG TDRRGPRARPI PQKSPWPWSGDAAKGQSGFLPVAAW AGQGRLPGGGIIVH
2527	A	2	614	PRVRLFTVITYFFVVGIGIPIFILIYELDSPLCWNEVF IGYGSALGSASFLTSFLGIWLFSCMEDIHMAFIGIF TTMTGMAMTAFASTTLMFLARVPFLFTIVPFVSVLRS

690
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion,=possible nucleotide insertion)
				MLSKVVRSTEQGTLFACIAFLETGGVTAVSTFNGIY SATVAWYPGFTFLLSAGLLLLPAISLCVVKCTSWNEG SYELLIQEESSEDASDRAC
2528	A	2	614	PRVRLFTVITYFFVVGIGIAPIFILYELDSPLCWNEVF IGYGSALGSASFLTSFLGIWLFSCMEDIHMAFIGIF TTMTGMAMTAFASSTLMMFLARVPFLFTIVPFSVLRS MLSKVVRSTEQGTLFACIAFLETGGVTAVSTFNGIY SATVAWYPGFTFLLSAGLLLLPAISLCVVKCTSWNEG SYELLIQEESSEDASDRAC
2529	A	1297	793	LGEPLGDLCELI PGDVQQLQMGEVHPGTGAQGSAAQS VAGEVQLTQLSHARQRPSCQGSQQLIALDLQHMDISRQ PRWQHVPVARQVQRAQQAQLAEGVAVHLWAGDAVVA EVELLQEVGGGKVFAANACDLVVQDHEGAHAARQATG HALQRVIVQVRRVQPLEAL*RVPSGLPRRVRAFMILH NQITGIGREDFATTYFLEELNLSYNRITSPQVHRDAF RKLRLRLSLDLSGNRLHMLPPGLPRNVHVLKVKNREL AALARGALAGMAQLRELYLTSNRLRSRALGPRAWVDL AHLQLLDIAGNQLTEIPEGLPESLEYLYLQNNKISAV PANAFDSTPNLKGIFLRFNKLAVGSVVDSAFRRLKHL QVLDIEGNLEFGDISKDRGRLGKEKEEEEEDEVEEEEE TR
2530	A	2	1671	LADGDMPLLLLLPLWGGSLQEKPVYELQVQKSVTVQ EGLCVLVPCSFSPWRSWYSSPPLYVYWFRDGEIPYY AEVVATNNPDRRVKPETQGRFRLLDGVQKKNCSLSIG DARMEDTGSYFFRVERGRDVKYSYQQNKLNLVETALI EKPDHIFLEPLESGRPTRLSCSLPGSCEAGPPLTFWS TGNALSPDPETTRSELTLTPRPEDHGTNLTCQMKR QGAQVTTERTVQLNVSYAPQTITIFRNGIALEILQNT SYLPVLEGQALRLLCDAPSNPPAHLWSFQGS PALNAT PISNTGILELRRVRSAAEEGGFTCRAQHPLGSLQIFLN LSVYSLPQLLGPSCSWEAEGLHCRCSFRARPAPSLCW RLEEKPLEGNSSQGSFKVNSSSAGPWANSSLI LHGGL SSDLKVSCAWNIYGSQSGSVLLLQGRSNLGTGVVPA ALGGAGVMALLCICLCIFFLIVKARRKQAAGRPEKM DDEDPIMGTTITSGSRKKPWPDSPGDQASPPGDAPPLE EQKELHYASLSFSEMKSREPKDQEAPSTTEYSEIKTS K
2531	A	2	1671	LADGDMPLLLLLPLWGGSLQEKPVYELQVQKSVTVQ EGLCVLVPCSFSPWRSWYSSPPLYVYWFRDGEIPYY AEVVATNNPDRRVKPETQGRFRLLDGVQKKNCSLSIG DARMEDTGSYFFRVERGRDVKYSYQQNKLNLVETALI EKPDHIFLEPLESGRPTRLSCSLPGSCEAGPPLTFWS TGNALSPDPETTRSELTLTPRPEDHGTNLTCQMKR QGAQVTTERTVQLNVSYAPQTITIFRNGIALEILQNT SYLPVLEGQALRLLCDAPSNPPAHLWSFQGS PALNAT PISNTGILELRRVRSAAEEGGFTCRAQHPLGSLQIFLN LSVYSLPQLLGPSCSWEAEGLHCRCSFRARPAPSLCW RLEEKPLEGNSSQGSFKVNSSSAGPWANSSLI LHGGL SSDLKVSCAWNIYGSQSGSVLLLQGRSNLGTGVVPA ALGGAGVMALLCICLCIFFLIVKARRKQAAGRPEKM

691
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				DDEDPIMGTTITSGSRKKPWPDS PGDQASPPGDAPPLE EQKELHYASLSFSEMKSREPDKQEAPSTTEYSEIKTS K
2532	A	51	674	QQAEEHLAAYSVSDSDSGKDPSMECCRRATPGTLLLF LAFLLLSSRTARSEEDRDGLWDAWGPWSECSRTC GGG ASYSLRRLSSKSCEGRNIRYRTCSNVDCPPEAGDFR AQQCSAHNDVKHHGQFYEWLPVSNPDNPNCSLKCAK GTTLVVELAPKVLDTGTRCYTESLDMCISGLCQVSADL FSFNLSRGFQCLCVNGLHSLTL
2533	A	239	577	GQPARVWSLDTMGTRLLPALFLVLLVLGFVQGTQQP QQDEMPSPFTFLTQVKESLSSYWESAKTAAQNLYEKTY LPAVDEKLRDLYSKSTAAMSTYTGI FTDQVLSVLKGE E
2534	A	239	577	GQPARVWSLDTMGTRLLPALFLVLLVLGFVQGTQQP QQDEMPSPFTFLTQVKESLSSYWESAKTAAQNLYEKTY LPAVDEKLRDLYSKSTAAMSTYTGI FTDQVLSVLKGE E
2535	A	103	318	MWRKHLSELLVLRDFLLAPRRRDSLTLTHMATLAQKPC GIEKQICFYVLFSLSLIFQHRLNSLKPRHLLRPDP*
2536	A	1	2374	MVSI SDLVICPPRHPKVLGLQGPPGLDSISDPSAGAG FLDWGEIGMPGPGRAGHQALCKCDCQCLEKTTTKAPG KMPKSTRSGPVRVRLADGPNRCAGRLECGMPDAGEQC VMTTGTSGRHCGLLGTGLWKGYTDLTIIPPGPGTPPQ ERTCQGDYHSGGTWTHSPLETTRPGSSSPAIRRLPA QMLLLPARPPHPRSSSPEAMDPPPPKAPFPKAEGPS STPSSAAGPRPRLGRHLLIDAN/GVYPYTYTVQLEE EPRGPPQREAPPGEPPGPRKGYSCPECARVFASPLRLQ SHRVSHSDLKPFTCGACGKAFAKRSSHLSRHRATHRAR AGPPHTCPLCPRRFQDAAELAQHSWGTPRGP LLAAAC NCEVARGRLESPPGERLLHGYGGREEEGGWGRAAGGL DRVEGFISSKAHHYLLIDTQGVPTVTLVTRSHRGSQG PVGLQARKVLQLPRVLKGLRVHVLQRHSITHSEVPQ DFAGSLDSFQTPGESLRLVFRALDTTQSSRISKAEPC LKEEPLSLGDLPYMHTTLCFCRKRRA SPGPGLQGA LAWPDWASPRALPVPSLSSTTRSPAAPLFAVPLSGRT TQAMAFDGIIFQGQSQRSAGLTTTSRFLACQRPLRLC AWWASRSRCTLRRPVGLRPGVHPRPRLVYRDLKPEN VMASGQPRDRPQPWFAPPPRPTFCGGCWTLTTPKEER CDRHQGAPGAPWRQREGAEAVGAVEERLGSEEAPGD AEREAAHPRPPRPTAFGVSSGLPELLVKRVVAQLQEL WTSSTAGGWSTQMOT
2537	A	241	957	MRSSLTMVGT LWAFSLVTA VTSSTSYFLPYWLFQSQ MGKPVSFSTFRRCNYPVRGEGHSLIMVEECGRYASFN AIPSLAWQMCTVVTGAGCALLLLVALAAVLGCCMEEL ISRMMGRCMGAAQFVGGLLISSGCALYPLGWNSPEIM QTCGNVSNQFQLGTCRLGWAYYCAGGGAAAAMLICTW LSCFAGRNPKPVILGGKHHEENHFLCYGAWPLPSTLE LRKEDRGGRATGKQVTP
2538	A	2817	1352	MAAAAAGAGSGPWAAQEKQFPALLSFFIYNPRFGPR EGQEENKILFYHPNEVEKNEKIRNVGLCEAIVQFTRT

692
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				FSPSKPAKSLHTQKNRQFFNEPEENFWMVMVVRNP I I EKQSKDGKPVIEYQEEELLDKVYSSVLRQCYSMYKLF NGTFLKAMEDGGVKLLKERLEKFFHRYLQTLHLQSCD LLDIFGGISFFPLDKMTYLKIQSFN\RMEESLNIVK YTAFLYNDQLIWSGLEQDDMRILYKYLTTSLFPRHI\ EPELAGRDSPIRAEMPGNLQHYGRFLTGPLNLNDPDA KCRFPKIFVNTDDTYEELHLI\VYKAMSAAVCFMIDA SVHPTLGF\CRRTGTASLGPQLHSGWASGHLVEQF*H QQGGCSGV*GKEPQFKFIYFNHNMNLAEKSTVHMRKTP SVSLTSVHPDLMKILGDINSDFTRVDEDEEIIIVKAMS DYWVVGKKSDRRELYVILNQKNANLIEVNEEVKKLCA TQFNNIFFLD
2539	A	171	347	NYSLSVYLVRLTAGTLLQKLRAKGIRNPDHSRALSE *HLSSLPHLIWIQVFLALQPS
2540	A	2	583	FPGRFRHNARRGFFFSHIGWLFVRKHRDVIKGRKL DVTDLLADPVVRIQRKYYKISVVLNCFVVPVTPVWYI WGESLWNSYFLASILRYTISLNI SWLVNSAAHMYGNR PYDKHISPRQNPLVALGAIGEGFHNYHHTFPFDYSAS EFGLNPNPTTWFI DFM CWLGLATDRKRATKPMIEARK ARTGDSSA
2541	A	1	1791	MTSGPQTSQPKEHLTNFKSDEQERVSSLAQSHTDNHR LHEPGLQEGIRAVPREDPQWNYQADSPRGPLDHHRRR ASGNSQWRQAKLIALTRALTAKGLRINIYTD SKYAF RILHHHVAWIWAERGFLPTQGSSIINATLIKTLKAAAL LPKEAGVIHCKGHQKASDPITQGNAYADKPIGFGLEK LLTFHLSQLQEYRGTKWREKSHRKVNHDENTS KLTSL NEEYTKNKTEYEEAQDAIVKEIVNISSGYVEPMQTLN DVLAQLDAVVSFAHVSNGAPVPYVRPAILEKGQGR I I LKASRHACVEVQDEIAFI PNDVYFEKDKQMFHIITGP NMGGKSTYIRQTGVIVLMAQIGCFVPCESAEVSI VDC ILARVGAGDSQLKGVSTFMAEMLETASILRSATKD SL I I IDELGRGTSTYDGFGLAWAISEYIATKIGAFCMFA THFHELTALANQIPTVNNLHVLTALTTEETLTMLYQVK KGVCDQSFGIHVAELANFPKHVIECAKQKALELEEFQ YIGESQGYDIMEPAAKKCYLEREQGEKIIQEFLSKVK QMPFTEMSEENITIKLKQLKAEVIAKNNSFVNEIISR IKVTT
2542	A	1	639	AGTARFVCQAEGIPSPKMSWLKNRKRKIHNSGR IKMYN SKLVINQIIPEDDAIYQCMAENSQGSILSRARLTVM SEDRPSAPYNVHAETMSSSAILLAWERPLYNSDKVIA YSVHYMKAEGLNNEEYQVVI GNDTTHYI I DDLEPASN YTFYIVAYMPMGASQMSDHVTQNTLEDGHTSVGLLQF AGGLLLTLVASVFPVPGDTTSEGCVTAK
2543	A	700	283	VPRLVSPLSNPAPKFYCVSFFYHMYGKHIGSLNLLVR SRNKGALDTHAWSLSGNKGNVWQQAHVPISPSGPFQI IFEGVRGPGYLGDI AIDDVTLKKGECPRKQTDPNKVV VMPGSGAPCQSSPQLWGPM AIFLLALQR
2544	A	2	673	NSRVEGQLCDLDPSAHFYGHCGEQLECR LDTGGDL SR GEVPEPLCACRSQSPLCGSDGHTYSQICRLQEAAARAR PDANLTVAHPGPCESGPQIVSHPYDTWNV TGQDVIFG

693
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				CEVFAYPMASIEWRKDGLDIQLPGDDPHISVQFRGGP QRFEVTGWLQIQAVRPSDEGTYRCLARNALGQVEAPA SLTVLTPDQLNSTGIPQLRSLNLVPEEEAESEENDDY Y
2545	A	195	635	IATMETKDQKKQKKNKSGPKAAKKKKRHLQDLQLGDE EDAWKRNPKAFQSAVWMARSFHRTQDLKTKKHHP VVDRTPLEPPPIVVVMGP/PKVGKSTLIQCLIRNFT RQKLTEIRGPVMIVSGKKLRLTIIDCGCDINMMIDLA
2546	A	167	691	MGVWVTLCTASACLTLLFWSQTPGKAFQIPCPPPHLS HWCLSPMQMDDGCARLCVLWTAWMRWRVLMCSCRVA TDLGIFLGVALGNEPLEMWPLTQNEECTVTGFLRDKL QYRSRLQYMKHYFPINYKIRVPYEGVFRIANVTRLRA QGSERELRYLGVLVSLSATESVHDELL
2547	A	1	337	RRFVSQETGNLYIAKVEKSDVGNVTCVVTNTVTNHKV LGPPTPLILRNDGVMGEYEPKIEVQFPETVPTAKGAT VKLECFALGNPVPTIIWRRADGKPIARKARRHKS RVG K
2548	A	2	462	EFQEAAKLYHTNYVRNSRAIGVLWAIFTICFAIVNVV CFIQPYWIGDGVDPQAGYFGLFHYCIGNGFSRELTC RGSFTDFSTLPSGAFKAASFFIGLSMMLIIACIICFT LFFFCNTATVYKICAWMQLTSAACLVLGCMIFPDGWD SDEVN
2549	A	418	768	AFTKHLKPRMEVKDCGAHNLEKGLTIFFHKGPPSSMY FRLCGPHEGRFFFL\IPPLHLLHLLFPLHFFYNFRDE ELSCTVVELKYTGNASALLILPDQDKMEEVEAMLLPE TFALCC
2550	A	2484	121	AIMTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTV TSCAVDQENQDPRRWVQKPPLNIQRPLVDSAGPRPKA RHOAETSQRLVGISQPRNPLEELRPSPRGQNVGPGPP AQTEAPGTIEFVADPAALATILSGEGVKSCHLGRQPS LAKRVLVRGSQGGTTQRVQGVASAYLAPRTPTHRLD PARASCFSRLEGPGPRGRTLCPQRLQALISPSGPSFH PSTRPSFQELRRETAGSSRTSVSQASGLLLETVPVQPA FSLPKGEREVVTHSDEGGVASLGLAQRVPLRENREMS HTRDSHDSHLMPSAPVAQPLPGHVPCPSPFGRQR VPSPGPPTLTSYSVLRRLTVQPKTRFTMPSTPRVQQ AQWLRGVSPQSCSEDPALPWEQVAVRLFDOESCIRSL EGSGKPPVATPSGPHSNRTPSLQEVKIQRIGILQQLL RQEVEGLVGGQCVPLNGGSSLDMLVELQPLLTEISR TLNATEHNSGTSHLPGLLKHSGLPKPCLPEECGEPQPCP PAEPGPPEAFCRSEPEIPEPSLQEQLEVPEPYPPAEP RPLESCCRSEPEIPESSRQEQLEVPEPCPPAEPPLPLE SYCRIEPEIPESSRQEQLEVPEPCPPAEPGLQPSTQ GQSGPPGPCPR\VELGASEPCTLEHRSLEPSLPP\CC SQWAPATTSLIFSSQ\HPLCASPPICSFQS\LRPPA\ GQAG/LSANLAPLEPLALKGA AFKSC\LTAIHCFHEA SSWTIECAF\YTSRAPP\SGPTRVCTNPVATLLEWQD ALCFIPVGSAA PQGSP
2551	A	356	1313	NCNLSVGSSCLSLASVWLARRMWTLRSPLTRSLVNM TSGPGGPAAAAGGRKENHQWYVCNREKLCESLQAVFV

694
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				QSYLDQGTQIFLNNSIEKSGWLFQIQLYHSFVSSVFSL FMSRTSINGLLGRGSMFVFS PDQFQRLKINPDWKTH RLLDLGAGDGEVTKIMSPHFEEIYATELSETMIWQLQ KKKYRVLGINWQNTGFQYDVISCLNLLDRCDQPLTL LKDIRSVLEPTRGRVILALVLPFHPYVENVGKWEKP SEILEIKGQNWEEQVNSLPEVFRKAGFVIEAFTRLPY LCEGDMYNDYYVLDDAVFVLKPV
2552	A	299	21	MGSSVLSIWILSPSIYPILSPLAMPCLSRDILIRVRR IQGAWPSEG TASSIRGWLT KL RMSSGKALEALYCIP GAAQHPGLGVTRVWSGRT*
2553	A	337	642	FAFPHYI KPYHLKRIHRAVL RGNLEKLYLLLTYYD ANKRDRKERTALHLACATGQPEMVHLLVSRRCENLC DREDRTPLIKAVQLRQEACATLLLLQNGA
2554	B	111	1520	PSIPAAVPQSAPPEPHREETVTATATSQVAQQPPAAA APGEQAVAGPAPSTVPSSTSKDRPVSQPSLVGSKEEP PPARSGSGGGSAAKEPQEERSQQQDDIEELETKAVGMS NDGRFLKFDIEIGRGSFKTVYKGLDTETTVEVAWCEL QDRKLT KSERQRFKEEAEMLKGLQHPNIVRFYDSWES TVKGKKCIVLVTELMTSGTLKTYLKRFKVMKIKVLR WCRQILKGLQFLHTRTPPIIHRDLKCDNIFITGPTGS VKIGDLGLATLKRASFASVIGTPEFMAPEMYEEKYD ESVDVYAFGMCMLMATSEYPYSECQNAAQIYRRVTS GVKPASF DKVAIPEVKEIIEGCIRQNKDERYSIKDLL NHAFFQEETGVRVELAEEDDGEKIAIKLWLRIEDIKK LKGKYKDNEAIEFSFDLERNVPEDVAQEMVESGYVCE GDHKTMAKAIKDRVSLIKRKREQRQL*
2555	B	111	1520	PSIPAAVPQSAPPEPHREETVTATATSQVAQQPPAAA APGEQAVAGPAPSTVPSSTSKDRPVSQPSLVGSKEEP PPARSGSGGGSAAKEPQEERSQQQDDIEELETKAVGMS NDGRFLKFDIEIGRGSFKTVYKGLDTETTVEVAWCEL QDRKLT KSERQRFKEEAEMLKGLQHPNIVRFYDSWES TVKGKKCIVLVTELMTSGTLKTYLKRFKVMKIKVLR WCRQILKGLQFLHTRTPPIIHRDLKCDNIFITGPTGS VKIGDLGLATLKRASFASVIGTPEFMAPEMYEEKYD ESVDVYAFGMCMLMATSEYPYSECQNAAQIYRRVTS GVKPASF DKVAIPEVKEIIEGCIRQNKDERYSIKDLL NHAFFQEETGVRVELAEEDDGEKIAIKLWLRIEDIKK LKGKYKDNEAIEFSFDLERNVPEDVAQEMVESGYVCE GDHKTMAKAIKDRVSLIKRKREQRQL*
2556	A	105	447	LIFCRVFEYLHSLHLPQEICLSLALFSRFTFCV IICE VDVWSVIFKVPFC SKRNKVAVHTMLYIQIFVSLFI*P QNWKQPKCPATVERINKMWYIHIV/EYYSANKR
2557	A	1	512	DEELPDLSVSRSSHLHWGIPVPGYDSQTIYVWLDAL VNYLTVIGYPNAEFKSWWPATSHIIGKDILKFHAIYW PAFLLGAGMSPQRICVHSHWTVCGQKMSKSLGNVVD PRTCLNRYTVDFRYFLLRQGVPNWDCDYDEKVVKL LNSELADALGGLLNRC TAKRIN
2558	A	1117	647	MILQVSGGPWTVALTALLMVLLISVVQSRATPENS VY QERQECYAFNGTQRVVDGLIYNREEYVHFDSAVGEFL AVMELGRPIGEYFNSQKDFMERKRAEVDKVC RHKYEL

695
TABLE 7

SEQ ID	Method	Predicted beginning nucleotide location of first amino acid residue of peptide sequence	Predicted ending nucleotide location of last amino acid residue of peptide sequence	Amino acid sequence (X=Unknown, *=Stop codon, /=possible nucleotide deletion, =possible nucleotide insertion)
				MEPLIRQRRGDVTITAVRGCWTTILSGYFLLKRGVVS GGCSWGSS*
2559	A	1027	254	STQRGGIKGVARAASLVGRRRAGTGMALLLCLVCLTA ALAHGCLHCHSNFSKKFSFYRHHVNFKSWWVGDI PVS GALLTDWSDDTMKELHLAI PAKITREKLDQVATAVYQ MMDQLYQGKMYFPGYFPNELRNIFREQVHLIQNAIIE SRIDCQHRCGIFQYETISCNNCTDSHVACFGYNCESS AQWKS AVQGLLN YINNWHKQDTS MRPRSSAFSWPGTH RAAPAFVLVLPALRCLEPPHLANLSLEDAA*CLKQH
2560	A	1027	254	STQRGGIKGVARAASLVGRRRAGTGMALLLCLVCLTA ALAHGCLHCHSNFSKKFSFYRHHVNFKSWWVGDI PVS GALLTDWSDDTMKELHLAI PAKITREKLDQVATAVYQ MMDQLYQGKMYFPGYFPNELRNIFREQVHLIQNAIIE SRIDCQHRCGIFQYETISCNNCTDSHVACFGYNCESS AQWKS AVQGLLN YINNWHKQDTS MRPRSSAFSWPGTH RAAPAFVLVLPALRCLEPPHLANLSLEDAA*CLKQH
2561	A	88	459	AGDHVSRNIPVATNNPVRVAVQEETRDRFHLLGDPQNK DCTLSIRD TRES DAGTYVFCVERGNMKWNYKYDQLSV NVTASQDLLSRYL E V P E S V T V Q E G L C V S V P C S V L Y P HYNWTASSPVYGS
2562	A	337	1129	AHLSARLSALILDEVAILPAPQNLSVLSTNMKHLMLW SPVIAPGETVYYSVEYQGEYESLYTSHIWI P S S W C S L TEGPECDVTDDITATVPYNLRVRATLGSQTS/CLEHP /VSIPLIETQPSLPDL/RMEITKDG FHLVIELEDLGP QFEFLVAYWRREP GAEEHV KMVRSGGIPVHLETMEPG AAYCVKAQTFVKAIGRYSAFSQTECV EVQGEAIPLVL ALFAFVGFM LILVVVPLFVWKMGRL LQ/YLLLPRGGS SQT P W K I T Q F
2563	A	1	359	ISGESIYWSQKPTPSSNASPWSEPAAVDVELTAYALL AQLTKPSLTQKEIAKATSIVAWLAKQRNAYGGFSSTQ DTVVALQALAKYATTAYVPSEEINLVVKSTENFQRTF NIQAVNRM
2564	A	150	299	MTFLILSIAPVLAVTGMIETAAMTGFANKDKQELKHA GKQLKLWRIYVL*

696
TABLE 8

SEQ ID NO:	Number of TM	TM range: scores
695	1	174-193:1980
696	1	49-73:2788
704	1	168-185:1769
711	1	4488-4504:2911
722	6	272-290:2864 328-351:1725 863-880:2348 1102-1128:3163 1137-1153:1708 1161-1180:2038
731	1	406-434:2245
732	1	579-607:2245
736	1	364-380:1936
740	1	302-321:2224
742	1	816-832:1758
756	1	1012-1028:1967
757	1	529-548:3334
758	1	533-552:3334
759	4	1014-1033:2221 1095-1113:2566 1171-1194:2506 1245-1265:2246
761	3	65-83:2205 117-136:2143 853-870:2248
773	3	73-88:2787 168-186:2328 340-360:2085
776	3	90-106:2479 212-232:2562 387-403:2183
781	1	115-132:1854
784	1	53-69:2130
795	3	433-453:1894 506-531:1812 606-622:2130
798	1	176-192:2849
804	1	231-248:3490
825	1	80-99:2954
826	1	194-213:2954
835	4	94-110:2105 145-161:1995 203-223:2483 366-385:1855
836	5	94-110:2105 145-161:2282 207-226:1712 427-442:1810 519-537:2682
838	1	530-547:3345
839	1	88-109:2169
842	1	149-175:1731
843	1	149-175:1731
846	1	300-316:1761
851	1	383-405:2659
852	1	379-401:2659
860	1	61-81:3175
866	2	62-81:1837 131-147:2154
871	1	50-68:2276
877	3	155-173:2724 426-442:2801 780-800:2540
883	3	192-214:1749 266-284:1879 425-444:2199
889	2	183-205:2141 304-320:2692
897	1	538-553:1709
898	1	725-740:1709
899	1	58-73:1930
901	1	102-121:2779

TABLE 8

SEQ ID NO:	Number of TM	TM range: scores
905	1	208-225:3345
906	1	116-133:2747
926	3	266-286:2107 431-450:2017 494-509:2005
927	1	307-329:2730
930	2	204-221:1978 259-275:1735
939	1	88-116:1861
950	3	343-368:2429 440-456:2054 498-513:2344
951	1	676-696:2381
952	1	79-95:2605
955	1	178-196:2063
958	1	394-414:2626
964	1	735-758:3292
968	1	84-99:2458
969	4	59-75:2180 119-134:2458 415-433:2785 501-522:2904
970	1	267-284:3132
975	3	192-208:2437 279-296:1885 392-409:2589
976	3	266-282:2437 353-370:1885 466-483:2589
992	1	1065-1083:1762
993	1	124-141:2188
996	1	450-474:2798
1003	1	313-334:2372
1018	5	71-95:2393 145-166:2340 187-204:1848 237-256:3231 297-318:1783
1023	1	239-257:2651
1024	1	377-395:1757
1025	1	339-357:1757
1032	3	192-214:1749 266-284:1879 425-444:2199
1039	2	152-168:2052 244-259:1761
1042	3	110-124:2032 198-214:1804 512-531:2204
1050	2	460-476:2094 570-590:2709
1055	1	306-332:2732
1062	2	82-97:2605 165-182:2300
1071	5	84-100:2101 214-230:2609 380-395:2074 456-478:1922 536-553:1999
1085	2	40-69:2283 99-120:1980
1094	4	93-108:2432 170-187:2464 205-220:2179 241-265:2052
1098	2	142-158:1937 197-216:2428
1099	1	550-567:3380
1110	1	105-127:2966
1117	2	225-240:1816 473-494:3219
1118	1	234-255:3219
1130	1	1245-1266:3138
1143	1	80-99:2954
1144	1	194-213:2954
1146	1	233-249:2778

698

TABLE 8

SEQ ID NO:	Number of TM	TM range: scores
1169	1	39-68:2097
1180	1	77-100:1932
1194	1	105-121:2609
1195	1	86-104:1835
1197	1	202-221:2761
1213	1	692-715:1701
1223	1	347-363:2829
1234	1	555-570:1891
1237	1	518-537:2980
1240	1	676-696:2930
1245	2	89-105:1701 156-172:2335
1247	1	856-879:3766
1249	1	211-237:3134
1251	2	82-99:2126 203-219:2134
1252	2	75-92:2355 196-212:2053
1264	3	189-206:2466 247-266:1853 321-336:1839
1265	1	580-604:2903
1266	1	580-604:2903
1274	1	56-70:2193
1275	1	719-739:2381
1279	1	155-175:2511
1284	3	89-105:1748 155-173:2433 350-366:2126
1289	1	471-489:2039
1290	1	195-212:1943
1292	1	241-263:2676
1293	1	241-263:2676
1306	1	610-625:2249
1310	1	201-221:1908
1313	1	201-217:2496
1315	1	59-75:2149
1316	1	59-75:2149
1319	4	200-217:2717 258-273:1781 295-318:2028 416-436:2373
1322	1	356-381:1996
1330	2	86-104:2471 167-190:2177
1337	1	194-209:1865
1341	2	144-165:2452 216-235:1700
1349	2	102-117:3056 174-195:2254
1363	1	435-452:2888
1364	1	235-254:3185
1368	1	114-134:1898

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
1	685	1369	1967	784_9546
2	686	1370	1968	784_9546
3	687	1371	1969	784_9546
4	688	1372	1970	784_9546
5	689	1373	1971	787_7048
6	690	1374	1972	784_2242
7	691	1375	1973	784_6005
8	692	1376	1974	788_2591
9	693			
10	694	1377	1975	789_2432
11	695			
12	696			
13	697	1378	1976	784_3765
14	698	1379	1977	784_6649
15	699			
16	700	1380	1978	784_6766
17	701	1381	1979	784_4050
18	702	1382	1980	787_10261
19	703	1383	1981	787_6018
20	704	1384	1982	784_6424
21	705	1385	1983	787_10201
22	706	1386	1984	785_2688
23	707	1387	1985	784_420
24	708	1388	1986	784_5130
25	709	1389	1987	789_1109
26	710	1390	1988	784_5141
27	711	1391	1989	784_2214
28	712	1392	1990	784_2214
29	713	1393	1991	784_5125
30	714	1394	1992	784_2076
31	715	1395	1993	784_2076
32	716	1396	1994	784_4128
33	717	1397	1995	787_2409
34	718	1398	1996	784_3232
35	719	1399	1997	784_10218
36	720	1400	1998	787_2961
37	721	1401	1999	784_1254
38	722	1402	2000	784_583
39	723	1403	2001	784_8056
40	724	1404	2002	784_3284
41	725	1405	2003	784_5767
42	726	1406	2004	784_1548
43	727	1407	2005	784_3819
44	728	1408	2006	784_582
45	729	1409	2007	784_1390
46	730	1410	2008	784_4142
47	731	1411	2009	785_3653
48	732	1412	2010	785_3653
49	733	1413	2011	785_3653
50	734	1414	2012	785_3653

700

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
51	735	1415	2013	784_31
52	736			
53	737	1416	2014	784_3092
54	738	1417	2015	784_382
55	739			
56	740	1418	2016	787_1538
57	741	1419	2017	785_226
58	742	1420	2018	784_2152
59	743	1421	2019	784_4772
60	744	1422	2020	784_3345
61	745	1423	2021	787_9691
62	746	1424	2022	787_9691
63	747	1425	2023	792_146
64	748	1426	2024	784_8428
65	749	1427	2025	789_1722
66	750			
67	751	1428	2026	784_767
68	752	1429	2027	784_4697
69	753	1430	2028	785_197
70	754	1431	2029	784_1601
71	755	1432	2030	792_7466
72	756	1433	2031	787_3014
73	757	1434	2032	784_1605
74	758	1435	2033	784_1605
75	759	1436	2034	784_6460
76	760	1437	2035	784_1606
77	761	1438	2036	784_1723
78	762	1439	2037	785_1480
79	763	1440	2038	784_9631
80	764	1441	2039	784_5962
81	765	1442	2040	784_5962
82	766	1443	2041	784_5962
83	767	1444	2042	784_7108
84	768	1445	2043	784_2392
85	769	1446	2044	784_4227
86	770	1447	2045	784_7743
87	771	1448	2046	784_561
88	772	1449	2047	790_421
89	773	1450	2048	789_6309
90	774	1451	2049	787_2543
91	775	1452	2050	784_3892
92	776	1453	2051	787_3685
93	777	1454	2052	784_8321
94	778	1455	2053	784_7951
95	779	1456	2054	784_4225
96	780	1457	2055	784_7169
97	781	1458	2056	784_5044
98	782	1459	2057	784_5670
99	783	1460	2058	784_2357
100	784	1461	2059	784_6637

701

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
101	785	1462	2060	784_3755
102	786	1463	2061	784_9196
103	787			
104	788	1464	2062	784_706
105	789	1465	2063	784_706
106	790			
107	791			
108	792	1466	2064	784_4289
109	793	1467	2065	784_7228
110	794	1468	2066	784_3033
111	795	1469	2067	784_6065
112	796	1470	2068	785_2882
113	797	1471	2069	785_2882
114	798	1472	2070	785_2882
115	799	1473	2071	784_7266
116	800	1474	2072	784_7453
117	801	1475	2073	784_7453
118	802	1476	2074	788_13662
119	803			
120	804	1477	2075	784_2527
121	805	1478	2076	784_2968
122	806	1479	2077	785_3195
123	807	1480	2078	785_3195
124	808	1481	2079	785_3195
125	809	1482	2080	790_14016
126	810	1483	2081	790_21053
127	811	1484	2082	787_9817
128	812	1485	2083	784_4047
129	813	1486	2084	784_4047
130	814	1487	2085	784_4047
131	815	1488	2086	787_9324
132	816	1489	2087	785_3086
133	817	1490	2088	785_3086
134	818	1491	2089	784_7345
135	819	1492	2090	784_8313
136	820	1493	2091	787_71
137	821	1494	2092	784_5644
138	822	1495	2093	790_16836
139	823	1496	2094	784_7226
140	824	1497	2095	784_1134
141	825	1498	2096	784_7001
142	826	1499	2097	784_7001
143	827	1500	2098	788_3086
144	828	1501	2099	787_1984
145	829	1502	2100	784_3145
146	830	1503	2101	784_3145
147	831	1504	2102	784_1806
148	832	1505	2103	784_1806
149	833	1506	2104	788_594
150	834	1507	2105	784_3693

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
151	835	1508	2106	785_531
152	836	1509	2107	785_531
153	837	1510	2108	784_7408
154	838	1511	2109	787_5951
155	839	1512	2110	790_632
156	840	1513	2111	792_5495
157	841	1514	2112	785_1317
158	842	1515	2113	784_8634
159	843	1516	2114	784_8634
160	844	1517	2115	784_4818
161	845	1518	2116	784_4818
162	846	1519	2117	785_793
163	847	1520	2118	784_1834
164	848	1521	2119	784_1834
165	849	1522	2120	784_295
166	850	1523	2121	787_2031
167	851	1524	2122	784_2673
168	852	1525	2123	784_2673
169	853	1526	2124	784_2673
170	854	1527	2125	784_3244
171	855	1528	2126	784_9676
172	856	1529	2127	784_7453
173	857	1530	2128	784_2939
174	858	1531	2129	784_2939
175	859	1532	2130	787_2042
176	860	1533	2131	787_2042
177	861	1534	2132	784_3037
178	862	1535	2133	787_8909
179	863	1536	2134	784_7563
180	864			
181	865	1537	2135	792_7045
182	866	1538	2136	790_1109
183	867	1539	2137	784_4483
184	868	1540	2138	784_4483
185	869	1541	2139	787_2061
186	870	1542	2140	784_5083
187	871			
188	872	1543	2141	785_571
189	873	1544	2142	784_2517
190	874			
191	875	1545	2143	784_2138
192	876	1546	2144	784_9072
193	877	1547	2145	787_9212
194	878	1548	2146	784_5182
195	879	1549	2147	784_5182
196	880	1550	2148	784_5182
197	881	1551	2149	788_11145
198	882	1552	2150	785_3208
199	883	1553	2151	785_2364
200	884	1554	2152	787_6120

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No., SEQ ID NO.) *
201	885			
202	886	1555	2153	785_2555
203	887	1556	2154	785_2555
204	888	1557	2155	788_5026
205	889	1558	2156	785_2399
206	890	1559	2157	785_316
207	891	1560	2158	784_8768
208	892	1561	2159	784_6600
209	893	1562	2160	785_3574
210	894	1563	2161	787_223
211	895	1564	2162	784_1272
212	896	1565	2163	784_1358
213	897	1566	2164	787_4447
214	898	1567	2165	787_4447
215	899	1568	2166	784_4287
216	900	1569	2167	784_7705
217	901	1570	2168	784_1214
218	902	1571	2169	784_3287
219	903	1572	2170	784_3287
220	904	1573	2171	784_3950
221	905	1574	2172	787_5951
222	906	1575	2173	788_8994
223	907	1576	2174	784_7827
224	908	1577	2175	784_952
225	909	1578	2176	784_952
226	910	1579	2177	784_952
227	911			
228	912	1580	2178	788_6394
229	913	1581	2179	784_6391
230	914	1582	2180	784_7670
231	915	1583	2181	784_4795
232	916	1584	2182	784_3004
233	917	1585	2183	784_3004
234	918	1586	2184	784_3004
235	919	1587	2185	790_1148
236	920	1588	2186	784_7696
237	921	1589	2187	787_7957
238	922	1590	2188	787_7957
239	923	1591	2189	787_7957
240	924	1592	2190	787_7957
241	925	1593	2191	787_7957
242	926	1594	2192	784_4718
243	927	1595	2193	785_3642
244	928	1596	2194	787_6699
245	929	1597	2195	784_6067
246	930			
247	931	1598	2196	784_8379
248	932			
249	933	1599	2197	784_6418
250	934			

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
251	935			
252	936	1600	2198	784_3080
253	937	1601	2199	792_3539
254	938	1602	2200	784_4948
255	939	1603	2201	787_4342
256	940	1604	2202	784_7815
257	941	1605	2203	784_5767
258	942	1606	2204	784_5767
259	943	1607	2205	784_5777
260	944	1608	2206	784_5777
261	945	1609	2207	784_5777
262	946	1610	2208	784_5777
263	947	1611	2209	784_4849
264	948			
265	949	1612	2210	787_6059
266	950			
267	951	1613	2211	784_3590
268	952	1614	2212	784_337
269	953	1615	2213	790_27506
270	954	1616	2214	784_6469
271	955	1617	2215	787_8139
272	956	1618	2216	784_3189
273	957	1619	2217	784_1459
274	958	1620	2218	790_11947
275	959	1621	2219	784_4007
276	960	1622	2220	784_4007
277	961	1623	2221	784_4007
278	962	1624	2222	784_4007
279	963			
280	964	1625	2223	784_1398
281	965	1626	2224	785_2523
282	966			
283	967	1627	2225	784_10126
284	968	1628	2226	785_3232
285	969	1629	2227	785_3232
286	970	1630	2228	784_9436
287	971	1631	2229	784_6743
288	972	1632	2230	789_4182
289	973	1633	2231	784_8857
290	974	1634	2232	784_1226
291	975	1635	2233	787_2898
292	976	1636	2234	787_2898
293	977	1637	2235	784_3743
294	978	1638	2236	790_1713
295	979	1639	2237	790_1713
296	980			
297	981	1640	2238	787_371
298	982	1641	2239	784_10083
299	983			
300	984	1642	2240	787_1611

705

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
301	985	1643	2241	787_1611
302	986	1644	2242	784_7755
303	987			
304	988			
305	989	1645	2243	784_264
306	990	1646	2244	784_9739
307	991	1647	2245	784_6525
308	992	1648	2246	784_4625
309	993	1649	2247	787_8999
310	994	1650	2248	787_2386
311	995	1651	2249	784_4743
312	996	1652	2250	784_6535
313	997	1653	2251	784_8245
314	998	1654	2252	784_4654
315	999	1655	2253	784_3551
316	1000	1656	2254	784_5827
317	1001	1657	2255	784_4984
318	1002	1658	2256	784_4984
319	1003	1659	2257	784_3145
320	1004	1660	2258	784_8058
321	1005	1661	2259	784_3657
322	1006	1662	2260	785_1191
323	1007	1663	2261	784_5580
324	1008	1664	2262	784_6281
325	1009	1665	2263	784_2185
326	1010	1666	2264	787_497
327	1011	1667	2265	784_4047
328	1012	1668	2266	784_8772
329	1013	1669	2267	791_3817
330	1014	1670	2268	791_3817
331	1015	1671	2269	784_8115
332	1016	1672	2270	784_3141
333	1017	1673	2271	784_3141
334	1018	1674	2272	787_1645
335	1019	1675	2273	785_256
336	1020	1676	2274	784_1733
337	1021	1677	2275	784_1858
338	1022	1678	2276	784_1858
339	1023	1679	2277	790_5163
340	1024	1680	2278	785_102
341	1025	1681	2279	785_102
342	1026	1682	2280	787_4041
343	1027	1683	2281	792_3856
344	1028	1684	2282	787_3012
345	1029	1685	2283	787_3012
346	1030	1686	2284	784_1108
347	1031	1687	2285	785_435
348	1032	1688	2286	785_2364
349	1033	1689	2287	784_2969
350	1034	1690	2288	784_7604

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of config nucleotide sequence	SEQ ID NO: of config peptide sequence	Identification of Priority Application that config nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
351	1035			
352	1036	1691	2289	787_3016
353	1037	1692	2290	784_2242
354	1038	1693	2291	790_2603
355	1039	1694	2292	787_6999
356	1040	1695	2293	784_3526
357	1041	1696	2294	784_6134
358	1042	1697	2295	784_5025
359	1043	1698	2296	784_2119
360	1044	1699	2297	787_2782
361	1045	1700	2298	784_10271
362	1046	1701	2299	785_2701
363	1047	1702	2300	784_9892
364	1048	1703	2301	785_1616
365	1049			
366	1050	1704	2302	785_366
367	1051	1705	2303	784_8058
368	1052			
369	1053	1706	2304	789_1756
370	1054	1707	2305	787_10036
371	1055	1708	2306	784_8381
372	1056	1709	2307	787_4467
373	1057	1710	2308	787_4467
374	1058	1711	2309	787_4467
375	1059	1712	2310	787_4467
376	1060	1713	2311	784_8234
377	1061	1714	2312	784_470
378	1062	1715	2313	784_8240
379	1063			
380	1064	1716	2314	784_9166
381	1065	1717	2315	784_7964
382	1066	1718	2316	790_21118
383	1067	1719	2317	784_6659
384	1068	1720	2318	784_8264
385	1069	1721	2319	787_2108
386	1070	1722	2320	784_4485
387	1071	1723	2321	784_4689
388	1072	1724	2322	785_1448
389	1073	1725	2323	785_3350
390	1074	1726	2324	784_4428
391	1075	1727	2325	787_5857
392	1076	1728	2326	784_8283
393	1077	1729	2327	784_8283
394	1078	1730	2328	784_1601
395	1079	1731	2329	784_1601
396	1080	1732	2330	784_1601
397	1081	1733	2331	784_1601
398	1082	1734	2332	784_1601
399	1083	1735	2333	785_3693
400	1084	1736	2334	788_8918

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
401	1085	1737	2335	787_757
402	1086	1738	2336	784_1907
403	1087	1739	2337	784_10178
404	1088	1740	2338	784_10178
405	1089	1741	2339	784_8535
406	1090	1742	2340	784_8535
407	1091	1743	2341	784_8535
408	1092	1744	2342	784_8301
409	1093	1745	2343	784_8301
410	1094	1746	2344	787_10129
411	1095			
412	1096	1747	2345	787_4498
413	1097	1748	2346	787_4498
414	1098	1749	2347	790_27173
415	1099	1750	2348	787_4500
416	1100	1751	2349	785_3699
417	1101	1752	2350	784_952
418	1102	1753	2351	784_952
419	1103	1754	2352	787_1871
420	1104	1755	2353	784_1835
421	1105	1756	2354	785_2845
422	1106	1757	2355	784_9214
423	1107	1758	2356	784_2232
424	1108	1759	2357	784_2232
425	1109	1760	2358	792_6149
426	1110			
427	1111	1761	2359	784_6702
428	1112	1762	2360	784_8354
429	1113			
430	1114			
431	1115	1763	2361	787_9215
432	1116			
433	1117	1764	2362	785_2878
434	1118	1765	2363	785_2878
435	1119	1766	2364	784_10026
436	1120	1767	2365	784_6265
437	1121	1768	2366	785_2731
438	1122	1769	2367	787_6236
439	1123	1770	2368	785_1252
440	1124			
441	1125			
442	1126	1771	2369	791_3415
443	1127	1772	2370	785_3334
444	1128	1773	2371	784_8215
445	1129	1774	2372	784_10074
446	1130	1775	2373	784_10074
447	1131	1776	2374	784_3863
448	1132			
449	1133	1777	2375	784_2811
450	1134	1778	2376	790_28311

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
451	1135	1779	2377	784_4221
452	1136	1780	2378	785_1480
453	1137	1781	2379	784_2520
454	1138	1782	2380	784_1312
455	1139	1783	2381	784_633
456	1140	1784	2382	785_590
457	1141	1785	2383	785_590
458	1142	1786	2384	790_12519
459	1143	1787	2385	784_7001
460	1144	1788	2386	784_7001
461	1145	1789	2387	788_5657
462	1146	1790	2388	784_4745
463	1147	1791	2389	787_6106
464	1148	1792	2390	787_2727
465	1149	1793	2391	784_3950
466	1150	1794	2392	790_10584
467	1151	1795	2393	784_2612
468	1152	1796	2394	787_2965
469	1153	1797	2395	787_2965
470	1154	1798	2396	787_8641
471	1155	1799	2397	785_3774
472	1156			
473	1157	1800	2398	784_8542
474	1158	1801	2399	784_8542
475	1159			
476	1160	1802	2400	790_13566
477	1161	1803	2401	785_410
478	1162			
479	1163	1804	2402	784_5054
480	1164			
481	1165	1805	2403	785_3036
482	1166	1806	2404	789_4683
483	1167			
484	1168	1807	2405	784_6816
485	1169	1808	2406	784_5981
486	1170	1809	2407	785_3078
487	1171	1810	2408	784_2586
488	1172	1811	2409	784_6539
489	1173	1812	2410	784_6539
490	1174	1813	2411	784_6539
491	1175	1814	2412	784_8016
492	1176	1815	2413	787_10370
493	1177	1816	2414	784_5450
494	1178	1817	2415	787_7533
495	1179	1818	2416	785_3119
496	1180	1819	2417	785_3120
497	1181	1820	2418	785_3122
498	1182	1821	2419	784_9756
499	1183	1822	2420	784_4843
500	1184	1823	2421	784_441

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
501	1185	1824	2422	784_1095
502	1186	1825	2423	784_1066
503	1187	1826	2424	785_206
504	1188	1827	2425	784_4128
505	1189	1828	2426	784_4128
506	1190	1829	2427	784_4128
507	1191	1830	2428	790_27336
508	1192			
509	1193	1831	2429	784_2678
510	1194	1832	2430	784_3456
511	1195			
512	1196	1833	2431	785_582
513	1197			
514	1198	1834	2432	789_4888
515	1199	1835	2433	789_4172
516	1200	1836	2434	784_9397
517	1201			
518	1202	1837	2435	784_1307
519	1203	1838	2436	789_5903
520	1204	1839	2437	784_9886
521	1205	1840	2438	784_2293
522	1206	1841	2439	784_5604
523	1207	1842	2440	784_7569
524	1208			
525	1209	1843	2441	784_9399
526	1210	1844	2442	784_5253
527	1211	1845	2443	784_8932
528	1212	1846	2444	784_7850
529	1213	1847	2445	787_10375
530	1214	1848	2446	792_2784
531	1215	1849	2447	784_2550
532	1216	1850	2448	784_3066
533	1217	1851	2449	785_2240
534	1218	1852	2450	785_76
535	1219	1853	2451	792_6297
536	1220			
537	1221	1854	2452	792_1062
538	1222	1855	2453	784_9474
539	1223			
540	1224			
541	1225	1856	2454	784_3898
542	1226	1857	2455	784_4445
543	1227	1858	2456	784_9615
544	1228	1859	2457	784_10126
545	1229	1860	2458	784_9880
546	1230			
547	1231	1861	2459	785_3774
548	1232	1862	2460	785_3774
549	1233	1863	2461	785_3774
550	1234	1864	2462	784_1315

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
551	1235			
552	1236	1865	2463	790_16605
553	1237	1866	2464	784_2311
554	1238	1867	2465	787_8252
555	1239	1868	2466	784_5605
556	1240	1869	2467	784_3824
557	1241			
558	1242	1870	2468	785_3563
559	1243	1871	2469	790_20271
560	1244			
561	1245			
562	1246	1872	2470	790_5164
563	1247	1873	2471	785_3680
564	1248	1874	2472	784_2988
565	1249	1875	2473	787_4774
566	1250			
567	1251	1876	2474	784_9364
568	1252	1877	2475	784_9364
569	1253	1878	2476	784_8765
570	1254			
571	1255	1879	2477	790_12841
572	1256	1880	2478	787_4398
573	1257	1881	2479	787_4398
574	1258			
575	1259			
576	1260	1882	2480	788_12600
577	1261	1883	2481	790_16405
578	1262	1884	2482	787_7025
579	1263			
580	1264	1885	2483	784_4168
581	1265	1886	2484	790_26483
582	1266	1887	2485	790_26483
583	1267			
584	1268	1888	2486	790_2440
585	1269			
586	1270	1889	2487	784_1755
587	1271			
588	1272	1890	2488	790_21097
589	1273			
590	1274	1891	2489	787_4393
591	1275	1892	2490	784_3590
592	1276	1893	2491	787_933
593	1277	1894	2492	790_8149
594	1278			
595	1279	1895	2493	787_6126
596	1280	1896	2494	785_3201
597	1281	1897	2495	784_360
598	1282	1898	2496	784_360
599	1283	1899	2497	784_360
600	1284	1900	2498	784_270

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
601	1285	1901	2499	784_5003
602	1286	1902	2500	784_6919
603	1287	1903	2501	790_27941
604	1288	1904	2502	790_19516
605	1289	1905	2503	785_1001
606	1290			
607	1291	1906	2504	784_1320
608	1292	1907	2505	785_3606
609	1293	1908	2506	785_3606
610	1294	1909	2507	784_8851
611	1295			
612	1296	1910	2508	792_4796
613	1297	1911	2509	787_1962
614	1298	1912	2510	787_1962
615	1299			
616	1300	1913	2511	791_4419
617	1301	1914	2512	784_287
618	1302	1915	2513	784_287
619	1303			
620	1304	1916	2514	784_4933
621	1305	1917	2515	784_4933
622	1306			
623	1307	1918	2516	784_1318
624	1308	1919	2517	784_3284
625	1309	1920	2518	784_3284
626	1310	1921	2519	784_915
627	1311	1922	2520	784_7261
628	1312	1923	2521	784_5106
629	1313	1924	2522	785_598
630	1314	1925	2523	787_4996
631	1315	1926	2524	785_1259
632	1316	1927	2525	785_1259
633	1317	1928	2526	792_4498
634	1318			
635	1319	1929	2527	784_4291
636	1320	1930	2528	784_4291
637	1321	1931	2529	784_7003
638	1322	1932	2530	784_7701
639	1323	1933	2531	784_7701
640	1324	1934	2532	784_2330
641	1325	1935	2533	789_6254
642	1326	1936	2534	789_6254
643	1327	1937	2535	785_2282
644	1328	1938	2536	790_23335
645	1329			
646	1330	1939	2537	785_2954
647	1331			
648	1332			
649	1333			
650	1334			

TABLE 9

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Identification of Priority Application that contig nucleotide sequence was filed (Attorney Docket No. SEQ ID NO.) *
651	1335	1940	2538	784_3290
652	1336	1941	2539	784_1408
653	1337	1942	2540	784_5274
654	1338			
655	1339			
656	1340			
657	1341	1943	2541	790_26963
658	1342			
659	1343	1944	2542	787_2980
660	1344	1945	2543	784_4818
661	1345	1946	2544	784_5145
662	1346	1947	2545	784_9169
663	1347	1948	2546	785_1586
664	1348	1949	2547	784_1600
665	1349	1950	2548	784_9629
666	1350	1951	2549	784_9248
667	1351	1952	2550	787_7062
668	1352	1953	2551	784_7286
669	1353			
670	1354	1954	2552	785_254
671	1355	1955	2553	784_8867
672	1356	1956	2554	784_7020
673	1357	1957	2555	784_7020
674	1358	1958	2556	788_1533
675	1359	1959	2557	787_2028
676	1360	1960	2558	785_2715
677	1361	1961	2559	784_6946
678	1362	1962	2560	784_6946
679	1363	1963	2561	784_935
680	1364	1964	2562	784_1103
681	1365			
682	1366	1965	2563	784_1601
683	1367	1966	2564	785_122
684	1368			

784_XXX = SEQ ID NO: XXX of Attorney Docket No. 784, US Serial No. 09/488,725 filed 01/21/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 784CIP, US Application Serial No. 09/552,317, filed April 25, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 784CIP3A/PCT, PCT Serial No. PCT/US00/35017 filed December 22, 2000, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

785_XXX = SEQ ID NO: XXX of Attorney Docket No. 785, US Serial No. 09/491,404 filed 01/25/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 785CIP3/PCT, PCT Serial No. PCT/US01/02623 filed January 25, 2001, which is incorporated herein by reference in its entirety, including Tables, and Sequence Listing.

TABLE 9

787_XXX = SEQ ID NO: XXX of Attorney Docket No. 787, US Serial No. 09/496,914 filed 02/03/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 787CIP, US Application Serial No. 09/560,875, filed April 27, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 787CIP3/PCT, PCT Serial No. PCT/US01/03800 filed February 5, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

788_XXX = SEQ ID NO: XXX of Attorney Docket No. 788, US Serial No. 09/515,126 filed 02/28/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 788CIP, US Application Serial No. 09/577,409, filed May 18, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 788CIP3/PCT, PCT Serial No. PCT/US01/04927 filed February 26, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

789_XXX = SEQ ID NO: XXX of Attorney Docket No. 789, US Serial No. 09/519,705 filed 03/07/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 789CIP, US Application Serial No. 09/574,454, filed May 19, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 789CIP3/PCT, PCT Serial No. PCT/US01/04941 filed March 5, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

790_XXX = SEQ ID NO: XXX of Attorney Docket No. 790, US Serial No. 09/540,217 filed 03/31/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 790CIP, US Application Serial No. 09/649,167, filed August 23, 2000, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 790CIP3/PCT, PCT Serial No. PCT/US01/08631 filed March 30, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

791_XXX = SEQ ID NO: XXX of Attorney Docket No. 791, US Serial No. 09/552,929 filed 04/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing Attorney Docket No. 791CIP, US Application Serial No. 09/770,160, filed January 26, 2001, which in turn is a parent application of continuation-in-part application bearing Attorney Docket No. 791CIP3/PCT, PCT Serial No. PCT/US01/8656 filed April 18, 2001, both of which are incorporated herein by reference in their entirety, including Tables, and Sequence Listing.

792_XXX = SEQ ID NO: XXX of Attorney Docket No. 792, US Serial No. 09/577,408 filed 05/18/2000, the entire disclosure of which, including sequence listing, is incorporated herein by reference. This application is the parent application of a continuation-in-part application bearing 792CIP3/PCT, PCT Serial No. PCT/US01/14827 filed May 16, 2001, which is incorporated herein by reference in its entirety, including Tables, and Sequence Listing.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-684.
2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.
3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 99% sequence identity with the polynucleotide of claim 1.
4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.
5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.
6. A vector comprising the polynucleotide of claim 1.
7. An expression vector comprising the polynucleotide of claim 1.
8. A host cell genetically engineered to comprise the polynucleotide of claim 1.
9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.
10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:
 - (a) a polypeptide encoded by any one of the polynucleotides of claim 1;
and
 - (b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-684.

11. A composition comprising the polypeptide of claim 10 and a carrier.
12. An antibody directed against the polypeptide of claim 10.
13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:
 - a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and
 - b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.
14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:
 - a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;
 - b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and
 - c) detecting said product and thereby the polynucleotide of claim 1 in the sample.
15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.
16. A method for detecting the polypeptide of claim 10 in a sample, comprising:
 - a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and
 - b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.
17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of any of the polynucleotides from SEQ ID NO: 1-684, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO: 685-1368.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising of at least one of SEQ ID NO: 1-684.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.